

# VU Research Portal

## Essays in Development Economics

Wiegand, Martin

2021

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Wiegand, M. (2021). *Essays in Development Economics*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam]. Rozenberg Publishers and Tinbergen Institute.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

Martin Wiegand

# Essays in Development Economics

# ESSAYS IN DEVELOPMENT ECONOMICS

MARTIN WIEGAND

ISBN 978-90-361-0664-1

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 786 of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

VRIJE UNIVERSITEIT

**Essays in Development Economics**

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy  
aan de Vrije Universiteit Amsterdam,  
op gezag van de rector magnificus  
prof.dr. C.M. van Praag,  
in het openbaar te verdedigen  
ten overstaan van de promotiecommissie  
van de School of Business and Economics  
op dinsdag 12 oktober 2021 om 13.45 uur  
in de aula van de universiteit,  
De Boelelaan 1105

door

**Martin Wiegand**

geboren te Leverkusen, Duitsland

promotoren:

prof.dr. C. Elbers  
prof.dr. M.P. Pradhan

promotiecommissieleden:

prof.dr. J. Aker  
dr. V. Alatas  
prof.dr. P.F. Lanjouw  
prof.dr. H. Oosterbeek  
prof.dr. E.T. Verhoef

*"I should be sorry if I only entertained them. I wish to make them better."*  
*Georg Friedrich Händel*





# Contents

<b>Acknowledgments</b>	<b>1</b>
<b>1 Introduction</b>	<b>5</b>
<b>2 Do Early-ending Conditional Cash Transfer Programs Discourage Continued School Enrollment?</b>	<b>11</b>
2.1 Literature and Potential Mechanisms . . . . .	14
2.2 Program and Data Description . . . . .	18
2.3 Identification Strategy . . . . .	24
2.4 Estimation . . . . .	29
2.5 Discussion and Possible Channels . . . . .	31
2.5.1 Discussion of Results . . . . .	31
2.5.2 Possible Channels . . . . .	32
2.5.3 Coefficient Stability . . . . .	36
2.6 Conclusion . . . . .	36
2.A Appendix . . . . .	38
2.A.1 List of Pre-treatment Characteristics . . . . .	38
2.A.2 Score Functions . . . . .	43
2.A.3 Repeated Cross-fitting . . . . .	45
2.A.4 Inference for the Model with Attrition . . . . .	46
2.A.5 Data Preparation and Hyperparameter Tuning . . . . .	52
2.A.6 Result Tables . . . . .	53
2.A.7 Omitted Variable Bias Estimation . . . . .	66
<b>3 Welfare Measurement and Poverty Targeting Based on Participatory Wealth Rankings</b>	<b>69</b>
3.1 Summary of the Field Experiment and Data Description . . . . .	72
3.2 Empirical Model . . . . .	77
3.2.1 Model of Satisfaction . . . . .	77

3.2.2	Within-village Targeting Accuracy and Welfare Measures . . . . .	77
3.2.3	Local Poverty Rate . . . . .	80
3.2.4	Total Targeting Accuracy . . . . .	80
3.3	Model Selection and Estimation . . . . .	81
3.3.1	Model Selection . . . . .	81
3.3.2	Estimation . . . . .	82
3.4	Results and Discussion . . . . .	85
3.5	Conclusion . . . . .	94
3.A	Appendix . . . . .	95
3.A.1	Imputation Procedure . . . . .	95
3.A.2	Household and Location Characteristics . . . . .	97
3.A.3	Additional Estimation Results . . . . .	98

<b>4</b>	<b>The Impact of Road Development on Household Welfare in Rural Papua</b>	
	<b>New Guinea</b>	<b>107</b>
4.1	Data and Context . . . . .	111
4.2	Estimation . . . . .	118
4.3	Results . . . . .	122
4.4	Conclusion . . . . .	131
4.A	Appendix . . . . .	132
4.A.1	Road Summary Statistics . . . . .	132
4.A.2	Impact Analysis of Surface Type and Condition . . . . .	134
4.A.3	Selective Migration . . . . .	138

<b>Bibliography</b>	<b>139</b>
---------------------	------------

<b>Summary</b>	<b>149</b>
----------------	------------

# Acknowledgments

Writing these lines marks the end of an important and eventful period of my life. On my journey toward this PhD thesis I have met many great people who made these years wonderful and stimulating—be it with technical advice and discussions, or with friendship, an open ear, and the occasional merrymaking. In the following paragraphs I want to express my gratitude to some of these people.

First and foremost, I must thank my two supervisors. My discussions with Chris Elbers have always been fruitful and inspiring. Chris's own curiosity and enthusiasm for problem solving never cease to amaze me. I remember once writing him to ask a technical question, but realized soon after that he had just left for vacation. That did not stop him from sending me a response a week later, on the day of his return. It contained a 10-pager outlining ideas and code run on simulated data, and the remark that he had a fun vacation thinking about my problem.

Menno Pradhan is an excellent advisor for writing and 'selling' my work, and he has helped me to assume a more pragmatic approach whenever I got side-tracked. We certainly bonded during our research trip to Papua New Guinea, such as when Menno's suitcase got stolen on the first day so that we had to share all but a toothbrush, or when I got dismally seasick during a diving excursion on our day off.

I am also thankful to Peter Lanjouw, who has helped me in shaping my third chapter with various long discussions and detailed comments on an early draft. For a time it almost seemed as if he was a third supervisor in disguise. Naturally, I am thankful to my coauthors as well, especially to Eric Koomen, whose prowess in GIS analysis and untiring attention to detail has been invaluable to the development of chapter four. The initiative of Christopher Edmonds made the same chapter possible in the first place.

I owe a lot to the members of the economics department, and especially those of the development group. The seminars, reading groups, the Monday morning meetings, and everyone's approachability have created an atmosphere that is welcoming and conducive to analytical and critical thinking. I also want to thank Trudi Heemskerk, Pienke Dekkers, Ester van den Bragt, and Judith van Kronenburg, whose reliable and patient support in

all matters non-research saved the day for me on various occasions.

My time at the Vrije Universiteit would not have been what it was without my fellow PhD students. I have tons of great memories of conferences, workshops, and other occasions with Lisette, Lukasz, and Zlata, or the ‘selfie group’. With my office mates David and David of the ‘funky room’, I had lengthy political discussions and shared lots of laughs and sarcasm. I cherish the positive vibes of both Philipp and Diana, two later additions to the development group. I also want to mention Karolina, Paul, Simas, Bo, Jurre, Zichen, Benjamin, Cindy, and Yeorim, who all made the VU a better place.

Going back further, I was lucky to have an awesome bunch of fellow MPhil students at the Tinbergen Institute. Uwe has been a beacon of calm and wisdom, beside being a great friend and awesome jazz pianist. With Travers, I cannot decide whether I respect him more for being a brilliant researcher or a tireless party animal. I’m grateful to Guilherme, Lenny, Stephanie, Ieva, and Simin, both for the countless hours of ~~suffering~~ productive work during group assignments as well as for all the fun and silliness. Thanks also to Alexandra, Swapnil, David, Sabina, Stephan, Pascal, Sandor, and Mehmet, for all the good times inside and out of TI.

Beyond those already mentioned, I forged some friendships that elevated my time in Amsterdam from good to terrific: Robin, with whom I first and truly discovered the heart and soul of Amsterdam’s nightlife; Magda, whose many talents include managing a sailing crew, improv acting at 5am, and being an awesome roommate; David, with whom I share my obsession for a children’s card game; Aydan, who has a gift for telling stories, and who always laughs about the same things as I; Gavin, who can fly a plane and once did a looping with me in it; and Angelica, who never loses her calm or curses people, unless threatened with a knife. This tight-knit group of friends, also known as the ‘vogels’, will surely persist whatever the future might bring. Celia, Dario, Silvia, Gabriele, Thomas, and Noor were also a part of this group, and I hope that our reunions will be many. Saskia is the person who showed me that it is possible to live in Amsterdam’s city center and not to be annoyed all the time.

I was fortunate to live together with some exceptional people in Amsterdam, a few of which I want to mention here: Quentin, who always reminded me that a PhD is not a sprint but a marathon; Antonin, who I have never seen in a bad mood; Natalie, who is so convincing at roommate interviews that she left ours with a key to the flat; Sjoerd, who knows the meanest drinking games and how to win them; Jean, who studies the brain and has grand plans for an on/off-switch for feeling hungry or tired; and Alessandro, who has so much overlap of interest with me that I suspect he might be a paid actor.

I also want to thank some of the people I already knew when I came to Amsterdam.

This includes my oldest friends from Düsseldorf: Alex and Arne, who know the hardships of doing a PhD first-hand and are always up for an adventurous trip; and Michail and Max, both with whom it is a delight to fall back into old habits. I also made lasting friendships during my studies in Bayreuth, particularly with Corinna, Tilman, and Harald, who all have made me a more reflected person and who always know when to add a healthy amount of absurdity to a conversation. I am extremely grateful to Sylvia, my former nanny and honorary family member. Until today I can be sure of her warmth and trust, and I strongly believe that without her I would have turned out a more boring person.

Lastly, and most importantly, I thank my parents. Their unconditional love and unwavering belief in me have proven the strongest support imaginable. They taught me early on that most anything can be accomplished with patience and perseverance—especially to my mother the phrase ‘giving up’ is entirely alien—virtues which certainly came in handy on the way to my PhD. It was also my mother who insisted on me diligently practicing the piano as a kid, which led me to experience early successes, build confidence, and started my lifelong love affair with this instrument. My father has never shied away from helping me wherever he could, be it with Latin coaching, moving houses, or driving for hours to hand in a printed version of my bachelor thesis minutes before the deadline. I miss the long conversations with him about life, love, and everything else. He did not live to see the end of my PhD, but it is safe to say that at my defense he would have been the proudest person in the room.



# Chapter 1

## Introduction

In recent times, it has become conventional wisdom that randomized controlled trials (RCTs) are the gold standard for causal inference. RCTs are no longer reserved to the exact sciences but have found their way into economics, political science, sociology, and psychology. In development economics, field experiments have led to better understanding of such diverse topics as microcredit provision, deworming programs, teacher salaries, and cash transfer programs, to name just a few. The 2019 Nobel prize awarded to Abhijit Banerjee, Esther Duflo, and Michael Kremer “for their experimental approach to alleviating global poverty” further exemplifies the trend. These successes notwithstanding, for many questions RCTs are infeasible due to financial, political, ethical, or technical constraints. This is where empirical microeconomics comes in. The field provides a collection of methods that emulate experimental conditions and are tailored for causal inference from observational data.

This dissertation consists of three independent studies in empirical development economics. They are connected in that they all try to answer questions that could not be tackled using experiments alone. At the same time, they make use of the experimental methodology as the guiding principle, and in some cases, of actual data from field experiments. In this introduction, I will briefly motivate and summarize the chapters. I will also illustrate for each study why it did not lend itself to experimentation, and touch on the respective econometric methods chosen for identification.

## Chapter 2

Chapter 2 is motivated, in part, by a personal anecdote. When I was in elementary school, perhaps at the age of 8, my enthusiasm for literature was lukewarm at best. My mother, being concerned with a potentially missed chance, decided to top up my pocket money for

every book I finished. Surely enough, this trick turned me into an avid reader. The effect lasted for perhaps a year, until it was decided—possibly after seeing the bewilderment of other parents with the practice—that I had grown out of needing such mundane incentives. This immediately turned me alliterate again. After all, I reasoned, what was the point of reading without payment? I sometimes wonder whether the temporary incentive scheme may have caused me to read even less overall in the years to come.

Conditional cash transfer (CCT) programs, where poor households receive benefits conditional on their children going to school, have some things in common with my mother's incentive scheme. CCTs tend to work well at achieving their primary goal (more children going to school). The rationale for conditionality is to nudge students (or their parents) towards behavior that is ultimately beneficial to them. But while being successful while they last, it is unclear what their aftereffects are if they are terminated prematurely, i.e., before the end of a student's potential school career. To follow up on this question, I look at data from PROGRESA, a Mexican CCT, in which education payments ended after the end of middle school.

I find that the transition probability to high school is in fact negatively impacted by the program. After ruling out competing explanations, I conclude that there are two likely reasons. First, the program could crowd out students' or parents' intrinsic motivation for seeking education. The payments thus turn going to school into a job, and once they stop there are no sufficient reasons left to continue. Second, the program may have anchored the perceived value of education to the payments. Them going to zero may be mistaken for a signal that school is simply not worth it after that point—if even the government is not willing to fund it further. I also find that the effect reverses for those students who have not been eligible to the program on grounds of not being counted as poor. They are more likely to pursue continued schooling if their peers received cash while going to middle school. It appears that PROGRESA encouraged students from non-poor families to use high school education to distinguish themselves from their disadvantaged peers.

The chapter is based on data from a large field experiment. However, the impact of PROGRESA on the conditional high school transition probability cannot be determined just based on a means comparison of middle school graduates from treatment and control villages. The reason is that the program itself almost certainly had an impact on the likelihood of completing middle school. Some middle school graduates in the treatment villages may indeed only have finished middle school because of the program. One might suspect that those students would have been less likely to go to high school than their peers for whom the monetary incentives were not decisive. In fact, for the conditional aftereffects of such an RCT, an ideal experiment does not exist, because the transition



probability of those leaving the sample is unobserved. To tackle this problem, I deploy a newly developed, semi-parametric identification method called *double machine learning* (Chernozhukov et al., 2018). It allows me to account for a very large number of pre-treatment variables, to balance the sample and thus overcome selection and attrition bias.

PROGRESA itself was deemed largely a success by academics and policy makers alike. It got scaled up and reintroduced under different names two times in Mexico (first as OPORTUNIDADES, then as PROSPERA), and was copied by many other countries. It thus came as a surprise to many when in early 2019 it was announced that PROSPERA was going to get canceled. The program had lost popular approval for several reasons, one of which may have been that it did a bad job of identifying poor households (Kidd and Athias, 2020). This highlights the importance of good targeting and how it affects public endorsement of social security programs, which is the subject of the next chapter.

## Chapter 3

Chapter 3 is on poverty targeting. Its starting point is again an existing field experiment, conducted in Indonesia by Alatas et al. (2012). It compares three different targeting methods to determine the beneficiaries of a one-time lump sum payment: a proxy means test (PMT), where household consumption is predicted based on household characteristics from a survey; a participatory wealth ranking (PWR), where representatives of each community rank households by their need for inclusion; and a hybrid method of the first two. The PWRs led to higher satisfaction than the other treatments and showed no evidence of favoring local elites or discriminating against minorities. But the authors also notice there are systematic differences between the PMT rankings and the PWRs, and that the outcomes of the PMT treatment (unsurprisingly) led to better targeting, when considering true household consumption as the targeting goal. I believe these findings make for a natural follow-up question: why use *consumption* as the benchmark measure to evaluate targeting effectiveness, and not a measure based on PWRs, that would much more reflects villagers' perceptions of poverty?

My first contribution for the chapter is to demonstrate how to construct such a measure, using the data collected by Alatas et al. (2012). The idea is to estimate the relationship of ranks from the PWR with various household welfare predictors. The resulting model can then be used to construct welfare scores for all households from the experiment, even those that were not in the communities that conducted PWRs. The scores define a welfare measure that makes inter-village comparisons possible—even though it is based on within-village rankings. It constitutes an alternative to per capita consumption

as targeting goal, and it can be used to quantify targeting effectiveness of various (actual and hypothetical) allocations.

Equipped with this new welfare measure, I explore the effects of using it as targeting benchmark on program satisfaction. I find that added precision in meeting the newly defined targeting goals does increase satisfaction. I do not find evidence that the participatory process itself makes people happier, after controlling for targeting accuracy. And lastly, it appears that targeting accuracy explains satisfaction outcomes better when it is based on the community rankings rather than consumption. With these insights, I argue that a PWR-based welfare score may be preferable to predicted consumption as targeting measure. This may be particularly useful in settings where conducting large-scale PWRs is not feasible or desirable. It is worth noting that when evaluating targeting methods, one can compute measures of their effectiveness against a targeting goal. But to assess targeting goals themselves, a separate independent yardstick is needed. This is a role that satisfaction fulfills nicely, in addition to being of inherent interest.

In distinction to chapter 2, there is an ideal experiment to answer the main questions here. One could simply add another treatment group to the existing trial, where the allocation of benefits follows PWR scores. But the questions of whether PWR scores are preferred to PMT scores and whether the ranking process matters at all for satisfaction can just as well be answered utilizing the existing setup. The chapter thus demonstrates a case where conducting or expanding on a costly experiment can be substituted with careful analysis.

Large-scale antipoverty programs such as the ones mentioned in chapters 2 and 3 typically require a certain stage of development to be effective. Besides a sufficient tax base and local government institutions, an important precondition is a functioning traffic infrastructure. Households cannot reliably receive or collect benefits if they cannot be reached via roads that are accessible all year round. Conditionalities for CCT programs, such as school attendance or health checks, cannot be fulfilled if schools and medical centers are simply out of reach. The last part of my thesis is set in rural Papua New Guinea (PNG)—a country where up until today only about two thirds of the rural population live within 2km of an all-season road (Slattery, Dornan and Lee, 2018).

## Chapter 4

Chapter 4 is joint work with Eric Koomen, Menno Pradhan, and Christopher Edmonds. In it, we study the role of road quality for the well-being of PNG's rural population. The positive impacts of roads on household consumption have already been demonstrated in numerous publications, including for PNG (Gibson and Rozelle, 2003). The chapter

expands on this by looking at distributional effects of roads. We are particularly interested to see whether better roads favor disadvantaged households relatively more or less. We also look at development outcomes on which roads have less obvious effects, such as formal employment and school enrollment.

We combine two rounds of geocoded household survey data from 1996 and from 2009/10, respectively, with corresponding road asset management data of the same time (which we obtained on the ground in PNG). Measuring the impact of roads is complicated, since in principle every stretch of a country's traffic network may be used by any household, with varying frequency. The heuristic we chose is to consider the surface type composition of the road connecting each household to the closest urban area, starting at the road segment closest to the household. Connections to towns are arguably the most important ones, enabling access to markets and most public services. The differences between sealed, gravel, and dirt roads are an important consideration given the relatively high rate of deterioration for sealed roads and the tradeoffs between road construction, upgrading, and maintenance resulting from constrained budgets for road works. We find that sealing roads leads to higher average consumption, housing quality, school enrollment, and reduces reliance on subsistence farming. Impacts seem to be higher for poor and remote households, indicating that road works can be considered pro-poor policy measures.

Roads do not get upgraded at random. Instead, factors like expected cost and economic gains as well as regional prosperity, governance, and geospatial conditions may all codetermine which roads become graveled, asphalted, or are being left to deteriorate. At the same time, the subject of roads does not lend itself to experimentation. A hypothetical RCT to test the impact of different surface types would always suffer from non-exclusivity of road use. Even disregarding this, it is hard to imagine that the political and economic cost of randomly sealing roads for an RCT would ever be justified, nor would the expectable insights reflect actual road policy. Nonetheless, the identification strategy in this chapter is guided by the principles of randomization and control. We estimate a correlated random effects model, making use of the repeated observations for the same routes, and thus effectively controlling for time-invariant factors. In addition, we control for regional trends and some carefully selected time-varying factors. The remaining correlation between road quality and outcomes can thus be treated as a causal relationship.

Our results show that upgrading the roads leading to the nearest town increases average household consumption, housing quality, and school enrollment, and reduces reliance on subsistence farming. An analysis by subgroups shows that the effects on consumption

and poverty are at least twice as high for households with a road distance of at least 30 km to the nearest town when comparing them to those living closer than 30 km. Furthermore, we apply a newly developed generalized quantile regression estimator (Powell, 2020) to look for effect heterogeneity along the distribution of consumption. The estimates suggest that upgrading dirt roads has a higher effect for the poorest households.

## Chapter 2

# Do Early-ending Conditional Cash Transfer Programs Discourage Continued School Enrollment?

Around 20 years after their first appearance, conditional cash transfer (CCT) programs for education—initiatives that provide financial incentives for poor households to send their children to school—have never been more popular. Praised for their potential to increase school enrollment while reducing poverty, they are now widespread in Latin America and gain quick traction in Africa and Asia. However, many programs do not cover the entirety of a student’s schooldays and instead stop with welfare payments after elementary school or middle school. Even the largest and most well-known CCT programs only started to cover high school students long after their introduction.<sup>1</sup>

In this chapter, I investigate how CCT programs can affect school enrollment after payments break off, using Mexico’s PROGRESA as a case study. The goal is to find out whether temporary financial incentives lead to a sustained increase in educational participation, or whether they have no or even negative effects on schooling after payments end. PROGRESA is particularly well suited for this research due to its program design in the first five years, since students were only covered until the end of middle school (approximately at age 15). The program was also accompanied by rigorous data collection for evaluation, and experimental conditions were created by the deference of the program

---

<sup>1</sup>This includes some of the largest CCTs at the time of writing (in terms of beneficiaries), namely Brazil’s BOLSA FAMÍLIA, the Philippines’ PANTAWID, and Colombia’s FAMILIAS EN ACCIÓN, as well as Mexico’s recently terminated PROSPERA (formerly known as PROGRESA). Further examples of currently running CCTs that do not cover high school are Indonesia’s PROGRAM KELUARGA HARAPAN, Ghana’s LIVELIHOOD EMPOWERMENT AGAINST POVERTY program, Pakistan’s PUNJAB FEMALE SCHOOL STIPEND PROGRAMME, Nigeria’s NATIONAL CASH TRANSFER PROGRAMME, and Burkina Faso’s NAHOURI CASH TRANSFERS PILOT PROJECT.

in some randomly chosen localities that would serve as the control group. These features make it possible to estimate the effect of PROGRESA on high school enrollment.

There is an extensive literature on the short and medium term effects of cash transfers on education outcomes, showing that CCT programs increase school enrollment while payments are in place and lead to more years of schooling overall (see review by Bastagli et al. 2016). However, to my knowledge there is no study on what a CCT program does to a student's likelihood to continue school once it is over.<sup>2</sup> The direction of these effects is not obvious: on the one hand, one might argue that earlier transfers free up resources, rendering continued schooling more likely. On the other hand, a number of theories from psychology and behavioral economics, such as loss aversion, motivation crowding, anchoring, and classroom peer effects, could explain why the probability to continue school might actually decrease due to earlier payments. Studying the aftereffects of CCT programs can thus shed some light on the interaction of financial incentives and the social norms and behavioral patterns that influence educational choice. In addition, the study is necessary to understand a CCT's full impact on the education distribution, and is highly relevant for the design of future programs: a policy maker with limited funds needs to worry less about early break-offs if CCT programs continue to have a positive effect on enrollment. If, on the other hand, it turns out that such programs actively discourage students from continued education, they may pose a trade-off between different levels of secondary education, and raise the question how such discouragement may be averted.

The effect of PROGRESA on high school enrollment can be expressed in two different quantities. The first one is an unconditional treatment effect. It measures the difference in probability of high school enrollment between treatment and control group, considering all adolescents from poor households who had finished primary school just before the program started (around the age of 12). The resulting number is easy to interpret, but is likely driven by the program impact on middle school enrollment. The second quantity is a conditional treatment effect, i.e., the effect on the probability to make the transition from middle school to high school. It is a more direct measure of program aftereffects, and thus receives the main focus of this chapter. The estimation is complicated by the fact that the education payments have likely changed the composition of middle school graduates between the treatment and control group after two years: presumably, some of the students in the treatment group would not have finished middle school in the

---

<sup>2</sup>A related problem is tackled by Buser et al. (2017), who study the impacts on child health of households dropping out of an unconditional cash transfer program in Ecuador, finding negative effects on weight and height. The methodological difference to my question is that I do not focus on the effect of PROGRESA ending (comparing those who lost eligibility with those who maintained it), but on its aftereffects (comparing participants after the end of the treatment with non-participants).

absence of payments. Consequently, these students have relatively fewer counterparts in the control group. To account for this, observations need to be weighted by the students' likelihood to graduate from middle school without the program, or equivalently, their likelihood to be in the treatment group given that they graduated from middle school.

The main result of the chapter is that having been paid for schooling in the past reduces the probability to continue once the payments stop by about 10 to 14 percentage points. It appears that making payments up to a point actively discourages students to stay in school afterwards. After testing and ruling out loss aversion and classroom peer effects as explanations, I conclude that this is likely caused by a shift in the perceived value of education. The negative effect does not spill over to adolescents from non-poor households in treatment locations, who were not eligible for the program. To the contrary, for these adolescents the program even had a positive high school enrollment effect, with an increase in high school graduation of between 15 and 19 percentage points. The program seems to raise the desire of non-poor students to distinguish themselves through more schooling.

The chapter's other contribution beside these findings is in demonstrating a way to estimate program aftereffects conditional on a composition change induced by the program itself. Despite the randomization of PROGRESA communities, the program payments can be expected to have led to differences in the distributions of middle school graduates between treatment and control group. It is worth noting that this imbalance is an inherent property of the evaluation of conditional aftereffects. There is no ideal experiment that might serve as a benchmark. Nonetheless, aftereffects can be relevant in various contexts, e.g. when studying withdrawal in medical trials or job search after a time-limited unemployment benefit program ends. The approach taken in this chapter is to make an assumption of unconfounded treatment, conditional on a large number of baseline characteristics. This is not done to correct for selection into treatment and control group (which is random), but for the decision to drop out as a result of group membership.

To obtain causal estimates, I employ a newly developed procedure to estimate treatment effects, called double machine learning (DML), by Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey and James Robins (2018). It is a doubly-robust estimation technique (see e.g. Bang and Robins 2005), i.e., it makes use of predictions of both propensity scores and outcomes, and is robust to misspecifications of either one of these. A variety of machine learning methods as well as sample-splitting are used to learn and predict the relationships of treatment status and outcomes with potential confounders. DML permits to capture complex functional relationships and to control for large sets of covariates—possibly containing more elements

than there are observations—without having to know either in advance. This approach relies on fewer assumptions than conventional propensity score methods, and makes results more credible in comparison. Beside correcting for selection bias, the method can be extended to correct for imbalances resulting from attrition as well. I develop the extension of the DML method in this chapter, using the assumption that missingness of outcomes is random conditional on observables.

The chapter proceeds as follows. In section 2.1, I summarize the related literature and offer a number of explanations for aftereffects of CCT programs on school enrollment. Section 2.2 sums up the relevant details about the program and the data used. In section 2.3, I explain the identification strategy. Section 2.4 contains details about the estimation procedure including the machine learning methods used, together with results. The findings are further expanded on in section 2.5, with an analysis of possible channels and a discussion of covariate stability. Section 2.6 concludes.

## 2.1 Literature and Potential Mechanisms

This chapter contributes to the literature on CCTs and school enrollment. For PROGRESA, it has been shown that adolescents of treatment families stay in school with higher probability than those in the control group, and that this effect is particularly pronounced for those in the age for middle school (Schultz 2004, Behrman, Sengupta and Todd 2005, and Behrman, Parker and Todd 2009). Todd and Wolpin (2006) and Attanasio, Meghir and Santiago (2012) evaluate the costs and benefits of alternative program specifications to the ones of PROGRESA using structural models of decision making. Both studies conclude that a shift of program resources from primary school age children to middle school age adolescents would have led to a higher increase of total completed years of schooling. Perhaps closest to this chapter, Behrman, Sengupta and Todd (2005) and Dubois, de Janvry and Sadoulet (2012) estimate transition probabilities of PROGRESA students from one grade to another. The study by Behrman, Sengupta and Todd (2005) is the only one that considers the probability to go to high school, however without controlling for compositional changes as a result of prior program exposure, and only for small subsets of the data. Dubois, de Janvry and Sadoulet (2012) aim to disentangle the effects of PROGRESA on grade repetition and continuation. They circumvent the selection problem by looking only at the first year of implementation, and find among other things that middle school students in the treatment group are more likely to repeat a grade. The authors speculate that this may reflect the incentive to stay in middle school longer due to the limited program coverage.



To my knowledge, no study so far has concentrated on the schooling impacts of PROGRESA—or any other CCT program—after payments stop. On the face of it, it may not be obvious why the decision to continue school should depend on having been paid to go to school before. In the remainder of this section, I offer a number of explanations of how this may come about, each one supported by theoretical or empirical literature.

*Easing financial constraints:* One of the goals of CCT programs is helping poor households to finance children’s education. If financial constraints are in fact the main driver of educational underinvestment, then easing these constraints by making cash payments should result in more schooling. For instance, a family may have saved just enough to allow their child to finish middle school. Giving transfer payments until that point may then enable the household to save more, which in turn might allow the child to go to high school. The study by de Janvry et al. (2006) supports the argument that CCT payments can help smoothing out spending on education. It finds that PROGRESA takes a safety net function, in that it protects children from the impacts of shocks on school enrollment. The smoothing of education spending may not only work intertemporally within households, but also between households. Angelucci et al. (2010) show that PROGRESA raises middle school enrollment only for children with large family networks, in which transfer payments go from better-off to worse-off family members to ensure their children’s school enrollment.

While CCT programs allow to save more money for future education and thus might facilitate high school enrollment, a number of insights from psychology and behavioral economics point in the opposite direction. In the following, I highlight four reasons why CCT programs could discourage further schooling after they end: loss aversion, motivation crowding, anchoring, and classroom peer effects.

*Loss aversion:* Loss aversion is a central feature of prospect theory (Kahneman and Tversky 1979, Tversky and Kahneman 1991). It means that relative to a psychological point of reference, losses loom larger than gains of equal size. For intertemporal choice problems, this means that people require a larger payment to postpone present consumption than the amount they are willing to pay to have future consumption now (Loewenstein 1988). In the context of CCT programs, the choice between working (more consumption now) and continuing school (more consumption later) may depend on whether going to school is framed as a loss or a forgone gain of current consumption. Since reference points are often derived from past levels of consumption, families who have received PROGRESA payments may frame the choice as having either less or the same current consumption as before. On the other hand, families who never received PROGRESA payments may perceive the choice as having either the same or higher levels of consumption, making them

more likely to choose more education. An early study documenting similar behavior is Weiss, Hall and Dong (1980), on the effect of the Seattle-Denver income maintenance experiment on education. The authors find that reducing the (relatively low) direct costs of schooling, by offering subsidies on schooling expenditure, led to a large increase in enrollment among young adults. At the same time, significantly reducing the opportunity cost of going to school, by increasing the income tax rate for low incomes, had no such effect. This finding is consistent with loss aversion, if the direct costs are perceived as losses while the opportunity costs are viewed as foregone gains (Thaler 1980).

*Motivation crowding:* Another relevant theory from behavioral economics is motivation crowding theory (Frey and Jegen 2001, Fehr and Falk 2002). It acknowledges that people's actions are often motivated by hope for social approval, a desire to be moral, or intrinsic interest. When monetary incentives are added, they can replace those motives. A famous example is given by Gneezy and Rustichini (2000), who show that introducing a fine for parents who are late to fetch their children from kindergarten makes them arrive even later. The explanation is that being late, which used to be the violation of an ethical norm before, is being reframed into a good that can be bought for a reasonable price. Importantly, removing the fine did not make the parents arrive earlier again. In the same way, PROGRESA may put a price tag on the moral obligation to let children go to school. In distinction to the experiment by Gneezy and Rustichini, the price of non-conformance is high enough to comply with the program. But quantifying the value of sending children to school may reduce the pressure to let them continue after the payments stop. The crowding out effect may also spill over from parents to students, who might view school as necessary labor rather than an opportunity to learn. The negative effect tangible rewards can have on students' intrinsic motivation to learn has been demonstrated in a number of psychological studies (see Deci, Koestner and Ryan 1999 for a meta-analysis).

*Anchoring:* If financial constraints were in fact the only reason for educational under-investment, there would be no reason to make transfer payments conditional on school attendance. Instead, an unconditional cash transfer could achieve the same result without the need to monitor compliance, and freed from the often raised criticism that CCT programs are paternalistic. One reason for conditionality is that children as well as parents may be poorly informed about the returns to education, or about the natural talent required to complete school (Fiszbein and Schady 2010). For instance, Nguyen (2008) shows that households in Madagascar lack information about returns to education but change decisions rationally when this information is updated. Jensen (2010) shows that eight-graders in the Dominican Republic massively underestimate the rate of return to secondary school. And Dizon-Ross (2018) finds that parents in Malawi hold inaccurate

beliefs about their children's ability, the more so when they have low education themselves, and that they misallocate resources to education accordingly. In this light, making cash transfer programs conditional is a way to nudge students into a higher level of educational attainment, thus overcoming not only financial obstacles but also bad decisions due to incomplete information. But in doing so, CCT programs also convey a signal about the value of education: if the government is willing to pay for it, it must be worth pursuing. Conversely, the drop in payments after middle school may suggest that subsidizing poor students to go to high school is not worth it—be it due to low marginal returns to schooling at this level, or because students from poor families are deemed unlikely to succeed there. This particular form of priming effect where a numerical reference point (the program payout) affects the assessment of an unknown value (the value of going to high school) is called anchoring (Tversky and Kahneman 1974). By first anchoring the value of schooling to the PROGRESA payouts and then reducing it to zero, the government could unintentionally make further education appear less desirable.

*Classroom peer effects:* If a CCT program works as intended, some students keep attending school who would not have done so in the absence of the program. Presumably, this leads to larger class sizes and a higher share of disadvantaged students, which may affect motivation and learning of the students who would have gone to school without the program. For instance, the literature on classroom peer effects suggests that higher shares of disadvantaged students lead to more misbehavior in class, lower teaching quality, and negative performance spillovers (e.g. Carrell and Hoekstra 2010, Lavy, Paserman and Schlosser 2012). Thus, by the end of middle school, some students may have lost their motivation or aptitude to continue with high school. However, the changes in the composition of students may well have heterogeneous effects on students with different background or ability, and may even increase motivation for some students (say, through increased competition for grades or out of a need to distinguish themselves from the disadvantaged students). Thus, classroom peer effects may affect high school enrollment in both directions.

All these channels are possible explanations of direct program effects. Most CCT programs only target the poorest households, but they may affect students from other households nonetheless. For the case of PROGRESA, such spillover effects have been documented: Angelucci and De Giorgi (2009) argue that due to inter-household risk-sharing, food consumption increases even for non-eligible households in PROGRESA treatment villages. And Attanasio, Meghir and Santiago (2012) find, for their sample of boys between 10 and 16, that school enrollment was substantially higher for the non-eligible adolescents in the treatment group than for those in the control group. If households within a village

share program resources, one would expect that spillover effects on high school enrollment take the same direction as for the eligible students. If enrollment rises as a result of increased savings, this effect might well spread across household networks to non-eligible students. Social norm changes, value signals, and classroom composition changes may affect non-eligible students differently than eligible ones. If any of these channels are among the main drivers for direct effects, the direction of spillover effects is a priori hard to predict.

## 2.2 Program and Data Description

The PROGRAMA DE EDUCACIÓN, SALUD Y ALIMENTACIÓN, short: PROGRESA, was a multi-component antipoverty program launched in 1997. Its original goal was to improve prior antipoverty programs in Mexico along a number of dimensions, such as increasing targeting efficiency and reducing administrative costs (Gantner 2009). At first, a limited number of rural localities were selected for inclusion. Localities had to have between 50 and 2,500 inhabitants, access to health and education services, and had to be considered highly deprived based on available census data. Households from the selected localities were then classified as poor or not poor based on baseline survey data. Only households classified as poor were eligible.

One declared target was to increase school attendance of adolescents from poor families. The education component of PROGRESA included cash transfers every two months to mothers of every child enrolled in grades 3 to 9 who attended at least 85 percent of classes. This includes the last four years of primary school (*primaria*) and all of middle school (*secundaria*) but not high school (*preparatoria* or *bachillerato*). Payments increased with the age of the child to adjust to the increasing opportunity costs of schooling due to higher child wages. However, according to Schultz (2004), these payments were still lower than the average value of full-time child labor. In 2001, the program was renamed OPORTUNIDADES and extended to urban areas, and the schooling grants were extended to include the high school level.

For evaluation purposes, localities were randomized into a treatment and a control group. Payments for eligible households in the treatment group started in May 1998, while payments for eligible households in the control group only started in December 1999. The evaluation sample includes 320 treatment localities and 186 control localities. Survey data was collected twice a year for all households of the evaluation set from 1997 to 2000. Two surveys were administered before the program started in the treatment group (in October 1997 and March 1998), three between the start of the program for the

treatment and the control group (in October 1998, May 1999, and November 1999), and three after the program had started for the control group (in May 2000, November 2000, and winter 2003). The surveys contain detailed information on each household, including demographics, expenditure, schooling and labor outcomes, attitudes toward education, and location characteristics.

To identify the effect that PROGRESA had on the transition to high school, I consider the cohort of students who could have started high school in the academic year of 2000/01. By the end of the term in July 2000, the eligible students in the treatment group had benefited from PROGRESA for more than two years. Those in the control group had only been exposed to the program for the last semester of middle school, when the decision to continue school afterwards had likely been made already.<sup>3</sup> Figure 2.2.1 depicts a timeline with all survey dates. The two arrows indicate how long the students from the cohort under discussion were exposed to the program, for the treatment and the control group, respectively.

I consider three outcome variables related to high school attendance. The first is whether the student went to high school at the time of the November 2000 survey, i.e., right after finishing middle school. The second variable is whether the student had ever been to high school by the time of the 2003 survey. The third variable is whether the student had completed high school in 2003 or was enrolled in the last grade, thus would supposedly have graduated by 2004. The last two variables are particularly useful to check the medium-term impact on high school enrollment. After all, it could be that any differences in high school enrollment at the end of middle school fade out after a while. This may happen, for instance, if parents and students do eventually overcome any behavioral biases induced by the program and start realizing their full education potential. Or it might be that those in the treatment group had actually formed expectations about a future inclusion of high school students in the program and thus simply postponed enrollment by a little.<sup>4</sup>

---

<sup>3</sup>One could also consider the students who were expected to start high school in the academic year 1999/2000. In July 2000, the eligible students in the treatment group of this older cohort had been exposed to the program for one year, and the students in the control group not at all. Unfortunately, however, for this cohort it is not possible to unambiguously determine the set of students having finished middle school in 1999, and whether these students continued to high school afterwards. This is due to the fact that some of the relevant questions do not appear in the corresponding survey rounds. Therefore, I do not consider this cohort.

<sup>4</sup>While this cannot be ruled out, it seems unlikely. The program itself had initially only been promised for three years, and its continuation after the general election in 2000 was uncertain. The decisions to continue and eventually extend the program to include high school students in 2001 were only made after an evaluation of PROGRESA by the International Food Policy Research Institute (IFPRI), based on the first few years of running the program (Schultz 2004, Skoufias 2005). Thus, given the information at the time, it would have been an unlikely bet for students to delay high school in hopes of its later inclusion.

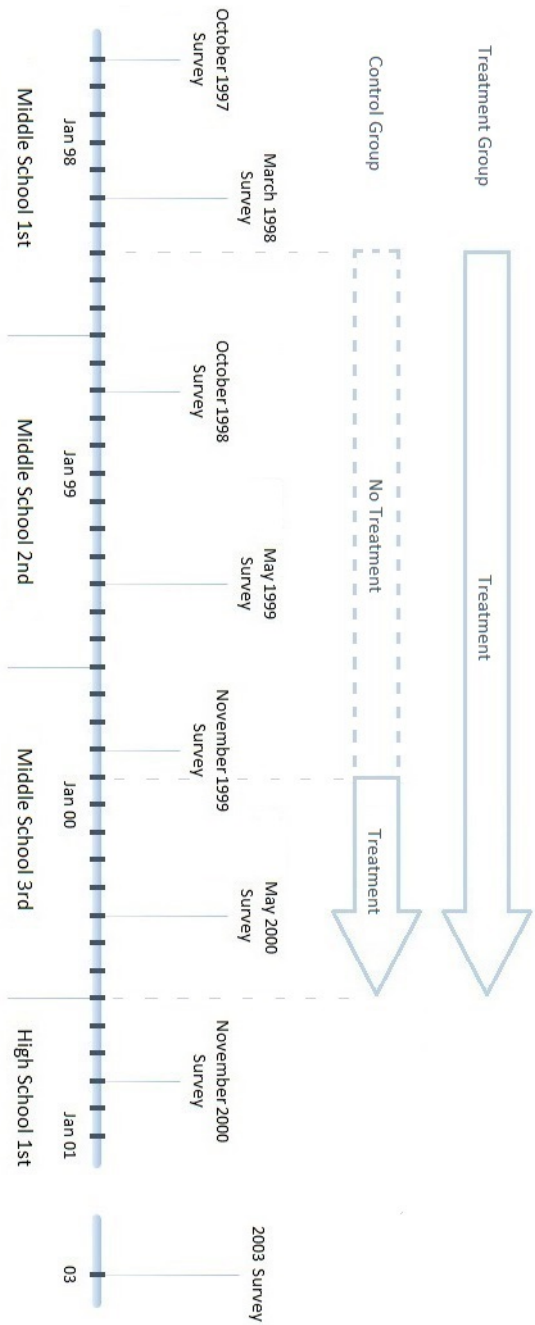


Figure 2.2.1: Timeline of relevant events around PROGRESA

In addition to high school outcomes I also consider two outcome variables related to middle school completion: whether a student graduated from middle school in academic year 1999/2000, and whether a student either graduated or was in the second last grade of middle school in academic year 1999/2000, thus would likely have graduated by 2001. The latter variable accounts for students who had a gap between primary and middle school, or who—voluntarily or not—repeated a grade in middle school. While these outcomes are not the main focus of this chapter, they serve as a way to verify prior results on the effectiveness of PROGRESA, and help to put the findings on high school enrollment into perspective. It is worth emphasizing that the average treatment effects on middle school completion and high school enrollment do not allow to make inferences on the average transition probability for middle school graduates, given the potential effect heterogeneity and differences between treatment and control group students at the end of middle school.

Tables 2.2.1 and 2.2.2 show descriptive statistics by eligibility status as well as treatment and control group. Table 2.2.1 summarizes the sample of all adolescents who had graduated from primary school in 1997 and who were between 11 and 14 years old at that time.<sup>5</sup> Thus, assuming a regular school career, these students could have started high school in 2000. I refer to this sample of students as the *unconditional sample*. Table 2.2.2 summarizes those adolescents of the same birth cohorts who finished the last year of middle school in 2000. I refer to this sample of students as the *conditional sample*, since it will be used to compute program treatment effects conditional on having graduated from middle school. Middle school completion is not included as a question in any of the questionnaires, but it can be constructed by taking all adolescents who reported the last year of middle school as their highest completed grade in the November 2000 survey, and who also reported being enrolled in school both in the November 1999 survey and in the May 2000 survey. There are two small caveats with this. First, the conditions do not rule out students who actually graduated from middle school before 2000 and then, in school year 1999/2000, attempted another grade of further education, which was not completed or simply not reported. Second, PROGRESA may have led some of the treated students to repeat the last grade on purpose to remain eligible to the program (see Dubois, de Janvry and Sadoulet 2012), and while it may not be wrong to include these students, it is conceivable that they influence the results significantly. So to exclude such cases, a further restriction is to consider only those students who reported the second year of middle

---

<sup>5</sup>The variable “had graduated from primary school in 1997” is in fact being constructed from multiple survey questions. The age restriction is being used as an additional fail safe against including people who completed primary school long before 1997. The interval 11 to 14 years is chosen to include students who started school at the regular age and repeated up to two grades. All the calculations in this chapter were also done for adolescents between 11 and 16 years as a robustness check. The results do not qualitatively differ from the ones presented in this chapter. They are available on request.

Table 2.2.1: Descriptive statistics of outcome variables, unconditional sample

	Treatment group				Control group				<i>t</i> -test	
	<i>N</i>	<i>M</i>	<i>SE</i>	<i>NA</i>	<i>N</i>	<i>M</i>	<i>SE</i>	<i>NA</i>	$\Delta$	<i>p</i>
<i>Adolescents from eligible (poor) households</i>										
Started high school in 2000	674	0.132	0.013	0.314	357	0.162	0.020	0.320	-0.030	0.327
Some high school by 2003	552	0.246	0.018	0.438	276	0.322	0.028	0.474	-0.076	0.104
Graduated or about to graduate from high school in 2003	552	0.178	0.016	0.438	276	0.228	0.025	0.474	-0.051	0.209
Graduated from middle school in 2000	713	0.418	0.018	0.274	373	0.434	0.026	0.290	-0.016	0.666
Graduated or about to graduate from middle school in 2000	713	0.663	0.018	0.274	373	0.576	0.026	0.290	0.087	0.020
<i>Adolescents from non-eligible (non-poor) households</i>										
Started high school in 2000	364	0.261	0.023	0.372	239	0.134	0.022	0.338	0.127	0.001
Some high school by 2003	331	0.347	0.026	0.429	194	0.258	0.031	0.463	0.090	0.091
Graduated or about to graduate from high school in 2003	331	0.296	0.025	0.429	194	0.180	0.028	0.463	0.116	0.012
Graduated from middle school in 2000	388	0.508	0.025	0.331	264	0.439	0.031	0.269	0.068	0.171
Graduated or about to graduate from middle school in 2000	388	0.644	0.024	0.331	264	0.602	0.030	0.269	0.042	0.345

*N* = number of observed outcomes, *M* = mean, *SE* = standard error, *NA* = share of adolescents with unobserved outcome,  $\Delta$  = difference in means, *p* = *p*-value for a two-sided *t*-test of equal means, clustered at the village level. The sample includes all adolescents who graduated from primary school and were of age 11-14 in 1997.



Table 2.2.2: Descriptive statistics of outcome variables, conditional sample

	Treatment group				Control group				<i>t</i> -test	
	<i>N</i>	<i>M</i>	<i>SE</i>	<i>NA</i>	<i>N</i>	<i>M</i>	<i>SE</i>	<i>NA</i>	$\Delta$	<i>p</i>
<i>Unrestricted conditional sample</i>										
<i>Adolescents from eligible (poor) households</i>										
Started high school in 2000	319	0.310	0.026	0.000	187	0.417	0.036	0.000	-0.107	0.097
Some high school by 2003	265	0.332	0.029	0.169	164	0.409	0.039	0.123	-0.076	0.241
Graduated or about to graduate from high school in 2003	265	0.268	0.027	0.169	164	0.341	0.037	0.123	-0.074	0.246
<i>Adolescents from non-eligible (non-poor) households</i>										
Started high school in 2000	195	0.415	0.035	0.000	152	0.355	0.039	0.000	0.060	0.326
Some high school by 2003	161	0.441	0.039	0.174	124	0.355	0.043	0.184	0.086	0.249
Graduated or about to graduate from high school in 2003	161	0.379	0.038	0.174	124	0.266	0.040	0.184	0.113	0.107
<i>Restricted conditional sample</i>										
<i>Adolescents from eligible (poor) households</i>										
Started high school in 2000	213	0.282	0.031	0.000	120	0.417	0.045	0.000	-0.135	0.065
Some high school by 2003	178	0.303	0.035	0.164	103	0.417	0.049	0.142	-0.114	0.147
Completed high school by 2004	178	0.253	0.033	0.164	103	0.340	0.047	0.142	-0.087	0.254
<i>Adolescents from non-eligible (non-poor) households</i>										
Started high school in 2000	123	0.447	0.045	0.000	100	0.340	0.048	0.000	0.107	0.165
Some high school by 2003	105	0.505	0.049	0.146	81	0.346	0.053	0.190	0.159	0.065
Completed high school by 2004	105	0.427	0.049	0.146	81	0.235	0.047	0.190	0.194	0.017

*N* = number of observed outcomes, *M* = mean, *SE* = standard error, *NA* = share of adolescents with unobserved outcome,  $\Delta$  = difference in means, *p* = *p*-value for a two-sided *t*-test of equal means, clustered at the village level. The unrestricted conditional sample includes all adolescents whose highest completed grade in November 2000 was the last grade of middle school, who were of age 14-17 then, and who were enrolled in school during the academic year 1999/2000. The restricted conditional sample has the further restriction that these adolescents had to be enrolled in the second last grade of middle school during the academic year 1999/2000.

school as their highest completed grade in the November 1999 survey. This restricted sample is defined more concisely, but comes at the cost of a loss of potentially relevant observations. Adolescents from eligible (poor) and ineligible (non-poor) households are regarded separately. The estimations of treatment effects are conducted for each of these two groups, to obtain direct program effects and spillover effects, respectively. The tables also report the share of missing observations ( $NA$ ) for each variable and experimental group.<sup>6</sup>

The descriptive statistics indicate that adolescents from eligible households in the treatment group went to high school with lower probability than those in the control group, both in the unconditional sample and the conditional samples. These differences cannot be interpreted as causal effects. It is to be expected that the composition of students who finish middle school differs between treatment and control group, and that missing outcomes are not missing completely at random. I construct a large number of exogenous characteristics to balance the two groups. They originate from the two pre-treatment surveys, and include such things as demographic and socioeconomic characteristics of the household, parents' level of education, parents' assessment of the student's ability and expectations about future educational outcomes, parents' assessment of teacher and school quality, village characteristics, average local education level, and travel times to a number of educational institutions. Using high-dimensional econometric techniques allows me to include a large number of potentially relevant characteristics without having to know in advance which of them are actually correlated with treatment status, high school enrollment, and outcomes missing. A list of all the considered characteristics is included in section 2.A.1 of the appendix.

## 2.3 Identification Strategy

The identification strategy is laid out here with the estimation of direct treatment effects in mind—with adolescents from eligible households as the population—but it works identically for the estimation of spillover effects. The problem of sample attrition is ignored for now, but is being discussed further below. For each of  $N$  students, let  $W = (Y, D, X)$ , where  $D$  is an indicator variable for living in a treatment locality and  $Y$  an outcome variable, e.g. an indicator for going to high school.  $Y(1)$  and  $Y(0)$  denote potential out-

---

<sup>6</sup>Missing outcomes are mostly a result of sample attrition. The panel also contains a few observations with inconsistent characteristics. These inconsistencies include a sex change, age discrepancies, and diminishing highest school degree. Those observations with inconsistencies between the pre-treatment surveys are dropped. Those with discrepancies only in one of the later surveys are deemed reliable with respect to their pre-treatment characteristics and only have outcomes set to missing.

comes, so that  $Y = DY(1) + (1 - D)Y(0)$ .  $X$  is a  $p$ -dimensional vector of exogenous control variables. All expectations are taken over the distribution of  $W$ .

In what follows, I distinguish between unconditional and conditional treatment effects, by which I mean the treatment effects for the correspondent samples, respectively. For the unconditional sample, the statistic of interest is the average treatment effect (ATE),

$$\text{ATE} = \mathbb{E}[Y(1) - Y(0)]. \quad (2.3.1)$$

For the conditional sample, the focus lies on those students who would have finished middle school even without PROGRESA. This is because the program itself has likely added some students to the pool of middle school graduates in the treatment group. These students do not have a counterpart in the control group, so that for them the treatment effect is not identifiable. On the other hand, it is inconceivable that a student who finishes middle school in the absence of payments would not have done so in their presence. To use the parlance of the literature on local average treatment effects: the analysis is concentrated on the *always-takers*, who by virtue of the experimental setup should be fully represented in both groups. It aims to leave out the *compliers*, as their counterfactual is not observed, as well as the *defiers*, who are nonexistent by assumption. One way to eliminate the compliers and defiers from the sample is to use a trimming technique (Imbens and Rubin, 2015), dropping students with no overlap in the distribution of covariates or propensity scores. This should eliminate both compliers and defiers from the sample for the ATE. Another way to exclude the compliers is to consider the average treatment effect of the non-treated (ATN),

$$\text{ATN} = \mathbb{E}[Y(1) - Y(0) | D = 0], \quad (2.3.2)$$

which is based on the distribution of students in the control group. As I am interested in the entire distribution of treatment and control group, the discussion focuses on the ATE. I do, however, compute estimates of both the ATE and the ATN and find that they are very close and statistically indistinguishable in all cases.

If the students under consideration were sampled into treatment and control group at random and if missing outcomes were missing completely at random, the ATE would simply be identified by the difference in sample means between treatment and control group, or average predictive effect (APE),

$$\text{APE} = \mathbb{E}[Y | D = 1] - \mathbb{E}[Y | D = 0]. \quad (2.3.3)$$

The APEs of the program are in fact equivalent to the differences in means ( $\Delta$ ) in Tables 2.2.1 and 2.2.2.

There are two reasons why the APE may not be an unbiased estimator of the ATE, despite the initial randomization of households. The first reason is sample selection, a concern only for the conditional sample: it may be that some students in the treatment group only finished middle school because of PROGRESA. These students would be comparatively less likely to continue to high school, and thus create the false impression that the program has a (more) negative effect on high school enrollment.

The second reason for possible bias is attrition. In the unconditional sample, around 33% of the adolescents identified in the two pre-treatment surveys have missing outcomes from the November 2000 survey, and around 45% of them have missing outcomes from the 2003 survey. In the unrestricted conditional sample, around 16% of the adolescents have missing outcomes from the 2003 survey. Attrition becomes a problem when it does not occur completely at random. For instance, it is conceivable that independently of their treatment status, the students who do not go to high school are more likely to drop out of the sample. Considering only those who stay would then lead the estimate of the ATE to be biased towards 0.

### Sample Selection Bias

I address these two concerns separately, starting with sample selection. For the estimation of conditional treatment effects, I rely on the assumption that treatment is independent of outcomes conditional on pre-treatment control variables  $X$ ,

$$Y(1), Y(0) \perp D \mid X. \quad (2.3.4)$$

Under this assumption, Rosenbaum and Rubin (1983) famously showed that it is sufficient to condition on the propensity score instead of the whole vector of controls. There are, however, some limitations commonly associated with this approach. The researcher needs to know exactly which variables to condition on, as well as the functional form of the probability model. Economic intuition may be helpful for model selection up to a point. But despite best efforts, seemingly relevant features may nonetheless lead to over-fitted propensity scores, while seemingly unrelated variables may hold a lot of predictive power through correlations with important unobserved features. In addition, the established methods require low model complexity for identification—i.e.,  $p \ll N$ —even in cases where a large number of confounders is plausible. Consequently, there is little insurance against misspecification of the probability model, which calls the unconfoundedness

assumption (2.3.4) and the propensity score method into question.

For this chapter, I use the specification and estimation strategy taken in Chernozhukov et al. (2018). To formalize the relationship between  $D$ ,  $Y$ , and  $X$ , consider the model

$$Y = g_0(D, X) + U, \quad \mathbb{E}[U | D, X] = 0, \quad (2.3.5)$$

$$D = m_0(X) + V, \quad \mathbb{E}[V | X] = 0. \quad (2.3.6)$$

This specification is quite general in that it allows for heterogeneous treatment effects and does not require  $D$  and  $X$  to be additively separable in the regression function  $g_0(D, X)$ .  $m_0(X)$  is the propensity score, i.e., the conditional probability to be in the treatment group. The subscript 0 indicates true parameters. The ATE is given by

$$\theta_0 := \mathbb{E}[g_0(1, X) - g_0(0, X)], \quad (2.3.7)$$

and the ATN by

$$\gamma_0 := \mathbb{E}[g_0(1, X) - g_0(0, X) | D = 0]. \quad (2.3.8)$$

Belloni, Chernozhukov and Hansen (2014) and Belloni et al. (2017) point out that in a high-dimensional parameter space, directly estimating equation (2.3.5) using sophisticated machine learning methods is ill-advised. While doing so may result in a great fit of  $Y$ , this approach neglects how treatment assignment is affected by covariates, potentially resulting in a large bias. One way to overcome this bias—and the approach taken in this chapter—is to use machine learning in conjunction with doubly-robust estimation, or double machine learning (Farrell 2015, Belloni et al. 2017, Chernozhukov et al. 2018).

The idea is to estimate the nuisance functions  $\eta_0 = (g_0(D, X), m_0(X))$  separately using machine learning methods.  $\theta_0$  and  $\gamma_0$  are then identified by plugging these estimates into a set of orthogonal moment conditions,  $\mathbb{E}[\psi_\theta(W; \theta_0, \eta_0)] = 0$  and  $\mathbb{E}[\psi_\gamma(W; \gamma_0, \eta_0)] = 0$ . The underlying score functions  $\psi_\theta(W; \theta, \eta)$  and  $\psi_\gamma(W; \gamma, \eta)$  for ATE and ATN are explained in section 2.A.2 in the appendix. Another crucial part of the DML method is cross-fitting, a sample-splitting technique to ensure that the same observations used to estimate nuisance functions  $g_0(D, X)$  and  $m_0(X)$  are not also used to make predictions thereof. This is done to prevent bias induced by overfitting, which is likely to occur for most machine learning techniques even after careful calibration of hyperparameters. The possibility to aggregate or choose the best out of multiple machine learning methods guarantees estimability for a wide range of data generating processes.<sup>7</sup> Details on the

---

<sup>7</sup>An alternative approach to deal with regularization bias is discussed in Athey, Imbens and Wager (2018). It does not require estimability of the propensity score, but in turn limits the complexity of the

cross-fitting procedure are given in section 2.A.3 of the appendix.

### Attrition Bias

Having discussed how the problem of nonrandom sample selection is approached, I now turn to nonrandom attrition. Let  $R$  be an indicator variable for attrition, taking the value 1 if  $Y$  is non-missing and 0 otherwise, and  $W = (Y, D, R, X)$ . I assume that outcomes are missing at random, meaning that attrition is independent of outcomes conditional on treatment status  $D$  and control variables  $X$ ,

$$Y(1), Y(0) \perp R \mid (D, X). \quad (2.3.9)$$

So, while attrition on its own may be predictive of outcomes, this predictive power comes entirely from observable variables. The approach is similar to the one taken in Behrman, Parker and Todd (2009) on the medium-term effects of PROGRESA, where attrition from the 2003 survey is also assumed to be random conditional on a (small) number of observables and treatment status. I propose the following extension of the model above to accommodate this assumption:

$$Y = g_0(D, X) + U, \quad E[U \mid D, R, X] = 0, \quad (2.3.10)$$

$$D = m_0(X) + V, \quad E[V \mid X] = 0, \quad (2.3.11)$$

$$R = r_0(D, X) + Z, \quad E[Z \mid D, X] = 0. \quad (2.3.12)$$

$r_0(D, X)$  is the conditional probability that student  $i$ 's outcome is observed.  $R$  does not enter the regression function  $g_0$ , but  $r_0(D, X)$  is needed to account for possible differences in the distribution of  $(D, X)$  between students with observed and unobserved outcomes. Just as for the model without attrition, ATE and ATN are identified by predicting nuisance functions  $\eta_0 = (g_0(D, X), m_0(X), r_0(D, X))$  and deploying the estimates in a corresponding orthogonal moment condition. The respective score functions are given in section 3.A.2 of the appendix.

As long as the nuisance functions are well approximated by any of the machine learning methods used, the resulting DML estimators  $\hat{\theta}_0$  and  $\hat{\gamma}_0$  are  $\sqrt{N}$ -consistent, approximately unbiased and asymptotically normally distributed. For the model without attrition, this is stated in Theorem 5.1 of Chernozhukov et al. (2018). An equivalent version of this theorem for the model with attrition is given in section 2.A.4 of the appendix of this chapter.

---

regression function by assuming strong sparsity.

## 2.4 Estimation

I estimate the ATE and the ATN using 10-fold cross-fitting with 100 repetitions. The sample is being split such that all students from the same location end up in the same fold. For the separate estimation of the nuisance functions  $\eta_0$ , I use six different machine learning methods. The first three machine learning techniques are regularized logistic regression techniques, namely the Lasso (with  $\ell_1$  penalty), Ridge (with  $\ell_2$  penalty), and elastic net (with both  $\ell_1$  and  $\ell_2$  penalty). Furthermore, I use two tree-based techniques—namely the random forest and extreme gradient boosting—and support vector machines (SVM).<sup>8</sup> In addition, I include a technique that combines the best machine learning methods for each nuisance function, i.e., the ones that produce the smallest out-of-sample mean squared error (or: Brier score). Before the actual estimation, the hyperparameters for each machine learning method are tuned to maximize out-of-sample predictive power. This is done via repeated cross-validation, using 10 folds and 10 repetitions. Details on hyperparameter tuning, data preparation for each method, and handling of missing feature values are given in section 2.A.5 of the appendix.

To exclude extreme values for the propensity score and to guarantee overlap between treatment and control group, I apply the trimming procedure developed in Crump et al. (2009) and Imbens and Rubin (2015). It produces an interval  $(\alpha, 1 - \alpha)$  such that all observations with propensity scores outside this interval are discarded. The number of trimmed observations varies by sample and cross fitting iteration, but it is 0 for the large majority of iterations for each sample. The maximum fraction of discarded observations in an iteration is 5.2% for the conditional samples and 0.3% for the unconditional sample.

For the unconditional sample, I compute the ATE for the three high school indicators and the two middle school indicators discussed above as outcomes.<sup>9</sup> For the conditional sample, I compute both the ATE and ATN for the three high school indicators. For the variance estimation of the DML estimators, it is necessary to account for clustering, since the treatment status of the PROGRESA experiment does not vary within villages. Since the number of observations per village varies substantially, cluster-robust standard errors may not be consistent, as is argued in Mackinnon and Webb (2017). This can be overcome by using a wild cluster bootstrap instead. I obtain standard errors in this way using

---

<sup>8</sup>In addition to these methods, I also considered neural networks with a single hidden layer, as well as different ensemble learners that would combine the aforementioned methods. Both the neural network and the ensemble methods turned out to be computationally expensive to tune, while showing relatively poor predictive performance. For that reason, I chose to leave them out eventually.

<sup>9</sup>In fact, since the unconditional sample is randomized, it would be sufficient for an unbiased estimate of the ATE to correct only for attrition. But accounting for covariates may increase precision even in the absence of sample selection, particularly if treatment and control group are not completely balanced, as is suggested by Behrman and Todd (1999) for the PROGRESA baseline data.

Table 2.4.1: ATE estimates of middle and high school education

Dependent variable	eligible (poor)			non-eligible (non-poor)		
	uncond.	cond.	restricted	uncond.	cond.	restricted
Started high school in 2000	-0.033 (0.030)	-0.125** (0.055)	-0.145** (0.069)	0.129*** (0.035)	0.066 (0.055)	0.093 (0.073)
Some high school by 2003	-0.085** (0.043)	-0.101* (0.059)	-0.126 (0.078)	0.100** (0.044)	0.109 (0.072)	0.162** (0.080)
Graduated or about to graduate from high school in 2003	-0.050 (0.035)	-0.100* (0.056)	-0.098 (0.076)	0.107*** (0.038)	0.149** (0.064)	0.193** (0.078)
Graduated from middle school by 2000	-0.029 (0.032)			0.069 (0.044)		
Graduated or about to graduate from middle school in 2000	0.088*** (0.032)			0.026 (0.037)		
Observations	1,507	506	333	941	347	223

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

The average treatment effect (ATE) together with standard errors (in parentheses) is estimated for the unconditional sample (adolescents who graduated from primary school in 1997) and for the unrestricted and restricted conditional samples (adolescents who graduated from middle school in 2000; in addition, the restricted sample has only those having been to the second-last grade of middle school in 1999), and separately for adolescents from eligible and non-eligible households. Treatment effects are estimated for five outcome variables: whether the student continued with high school in school year 2000/01, whether the student had attended some high school by the end of 2003, whether the student had finished or was about to finish high school by the end of 2003, whether the student had finished middle school by the end of 2000, and whether the student had finished or was in the last grade of middle school by the end of 2000.

Point estimates are obtained via the respective orthogonal moment conditions (see appendix section 2.A.2). Standard errors are obtained via wild cluster bootstrap. Results are based on 10-fold cross-fitting, using 100 sample splits. Estimates from the sample splits are combined using the median method, based on equations (2.A.6) and (2.A.7). Only results from combining the best-fitting machine learning methods for each nuisance function are reported here. Estimates of ATE and ATN for all the machine learning methods are reported in section 2.A.7 of the appendix.



100,000 bootstrap replications and the distribution for the bootstrap multiplier suggested by Mammen (1993). The ATE results from combining the nuisance function estimates of the respective best-fitting machine learning methods (i.e., my preferred specification) are reported in Table 2.4.1. More detailed results, with estimates for each of the machine learning methods used and including the ATN, are depicted in Tables 2.A.2 to 2.A.9 in section 2.A.6 of the appendix.

## 2.5 Discussion and Possible Channels

### 2.5.1 Discussion of Results

I start by discussing the eligible students, since they are the main focus of this chapter. Looking at the first column of Table 2.4.1, it appears that the program did not have a statistically or economically significant effect on timely high school enrollment for the eligible student population. However, it seems that by 2003, the program made it less likely by about 8.5 percentage points for students to have enrolled in high school at some point. Looking at the third outcome, it is unclear whether this translated into lower high school graduation rates for treated students. It is a somewhat unexpected result that eligible students in the treatment group who had just completed primary school when the program started would not have *higher* eventual high school continuation rates. This is especially so since PROGRESA seems to have had a positive effect on middle school completion: the probability to have graduated from or to be in the last grade of middle school in 2000 went up by about 8.8 percentage points.<sup>10</sup> Therefore, it must be that the program had a negative effect on the continuation decisions of middle school graduates. This hypothesis is confirmed when looking at the second and third columns. The ATE for high school in 2000 is -12.5 and -14.5 percentage points for the unrestricted and restricted sample, respectively. The effect is slightly smaller when looking at high school participation until 2003, with -10.1 and -12.6 percentage points, and at high school graduation or near graduation by 2003, with -10.0 and -9.8 percentage points.

For the non-eligible students, the fourth column of Table 2.4.1 shows that being in the treatment group increased high school enrollment and graduation by up to 13 percentage points. This result is partly driven by the program's known spillover effect of middle

---

<sup>10</sup>However, the effect is close to zero when only considering middle school graduation by 2000. The difference between the effects on graduation by 2000 and 2001 could arise because some adolescents may have enrolled in middle school in response to the program after it was launched in 1998, while the program did not affect the completion rate of those already enrolled at that time. Another explanation is that some students in the treatment group purposefully repeated a grade to receive payments longer, as is conjectured by Dubois, de Janvry and Sadoulet (2012).

school graduation on non-eligible students (Attanasio, Meghir and Santiago, 2012), which is also confirmed here. Looking at the fifth and sixth column, it seems that it is also driven by the program's direct effect on middle school graduates: the ATE for all high school outcomes is positive, and it is statistically significant for high school graduation and (for the restricted sample) for having done some high school by 2003. This result is unexpected, since it implies a sort of reverse spillover effect: those who saw their peers getting paid became more likely to continue schooling, while those actually getting paid lost interest. There are nonetheless ways to rationalize this finding. One explanation may be that an influx of poor students in middle school increases the need of non-poor students and their parents to distinguish themselves through further education. It is also conceivable that seeing their peers getting paid triggers the ineligible students' will to demonstrate their capability despite being at a relative disadvantage. Lastly, classroom peer effects could be responsible. An increase in eligible low ability students in middle school as a result of the program may reduce learning outcomes for other low ability students, while it may strengthen the relative position of high ability students (of which many may be non-eligible) and thus heighten their self-esteem and motivation. This latter hypothesis is, however, not supported by the data, as is shown in the following subsection.

### 2.5.2 Possible Channels

The precise channels for the observed treatment effects are impossible to learn with certainty from the data at hand. Nonetheless, I can examine two of the possible causes mentioned in section 2.1, namely classroom peer effects and loss aversion, and check whether they constitute credible explanations for the observed effects. Starting with classroom peer effects: if it is true that PROGRESA worsens the pool of middle school students, one would expect this to show in measures of performance. Unfortunately, such measures are not available until the 2003 survey. However, parents' assessments are available from the pre-treatment survey of March 1998. I consider three binary variables: whether (according to the parents) the student is good at school, whether the student is apt enough to go to high school or further, and whether the student is apt enough to go to university. While these assessments are certainly very noisy signals of a student's ability, there is no obvious reason to believe they should not at least convey some information thereon. Table 2.5.1 shows the differences in averages of the assessment variables between the treatment and control group, both for eligible and non-eligible students. Judging by this table, the hypothesis that the eligible students who finished middle school in the treatment group should on average be less apt than those in the control group is not supported. If anything, the numbers suggest the opposite. Thus, the theory of negative classroom peer

effects is not supported by the data.

Next, in order to examine loss aversion as a possible explanation, I check whether financial concerns are responsible for the differences in high school enrollment between treatment and control group. To that end, I look at a question asking for the reasons why students did not go to school in the November 2000 survey. Of the possible answers to this question, three point to financial constraints, namely: (1) there is not enough money to send the student to school, (2) the student is needed for work, and (3) the student is needed at home. I lump these together in an indicator variable that is 1 if one of these three reasons were given and 0 otherwise. In the same fashion, I create another indicator that lumps together all the other reasons why a student might not go to high school. The most frequent reasons here include: (1) the student does not like school, (2) the student is in vocational training<sup>11</sup>, (3) the student is already grown up, and (4) the school is too far away. Yet another variable related to financial constraints may be actual household expenditures. Therefore, I also construct a variable of monthly per capita expenditure, using prices and quantities of goods as indicated in the section on consumption and expenditure from the November 2000 survey.

I use these three variables—student does not go to high school due to financial reasons, student does not go to high school for other than financial reasons, and monthly per capita expenditure—as outcomes and apply the DML model on the conditional samples. This way it should be possible to see whether or not the program effects can be explained through intermediate effects on household finances. In particular, if loss aversion is the main driver of the negative treatment effect for the eligible students, this might show in a higher share of students not going to school for money reasons and higher per capita expenditures in the treatment group. Table 2.5.2 sums up the results, with more detailed outputs in Tables 2.A.10 to 2.A.13 in section 2.A.6 of the appendix.

The results show that not going to high school due to financial constraints as well as expenditure seem to be nearly unaffected by the program. On the other hand, the program does seem to disincline eligible students from high school for other than financial reasons, whereas it has the opposite effect on non-eligible students. This result indicates that loss aversion may not be the main explanation for the negative treatment effect on eligible students. Instead, it seems more likely that the program impacts enrollment through other factors such as social norms or students' motivation.

---

<sup>11</sup>One could argue that vocational training is a choice that promises security and may thus be counted as a signal for financial constraints. Only 3% of the students in the unrestricted sample had chosen this track, and counting them to the other group does not significantly change the results.

Table 2.5.1: Pre-treatment ability assessments by parents of middle school graduates

	Treatment group			Control group			<i>t</i> -test	
	<i>N</i>	<i>M</i>	<i>SE</i>	<i>N</i>	<i>M</i>	<i>SE</i>	$\Delta$	<i>p</i>
<i>Adolescents from eligible (poor) households</i>								
Student is good at school	264	0.614	0.030	156	0.577	0.040	0.037	0.496
Student can make it at least to high school	315	0.352	0.027	181	0.287	0.034	0.065	0.223
Student can make it to university	355	0.117	0.018	181	0.061	0.018	0.057	0.062
<i>Adolescents from non-eligible (non-poor) households</i>								
Student is good at school	160	0.613	0.039	132	0.621	0.042	-0.009	0.889
Student can make it at least to high school	187	0.380	0.036	151	0.411	0.040	-0.031	0.640
Student can make it to university	187	0.187	0.029	151	0.172	0.031	0.015	0.760

*N* = number of observed outcomes, *M* = mean, *SE* = standard error,  $\Delta$  = difference in means, *p* = *p*-value for a two-sided *t*-test of equal means, obtained via wild cluster bootstrap. Based on the unrestricted conditional sample.

Table 2.5.2: ATE estimates of PROGRESA on reasons not to continue school and household expenditure

Dependent variable	eligible (poor)		non-eligible (non-poor)	
	unre- stricted	restricted	unre- stricted	restricted
Not going to high school due to financial constraints	0.035 (0.053)	0.031 (0.064)	0.002 (0.057)	-0.010 (0.072)
Not going to high school for other than financial reasons	0.073* (0.042)	0.113** (0.051)	-0.087* (0.046)	-0.085 (0.060)
Log monthly per capita expenditure	0.059 (0.063)	0.019 (0.077)	-0.079 (0.081)	-0.113 (0.101)
Observations	506	333	347	223

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

The average treatment effect (ATE) together with standard errors (in parentheses) is estimated for the unrestricted and restricted conditional samples, and separately for students from eligible and non-eligible households. The unrestricted sample consists of all adolescents who completed middle school in 2000 and were of age 14-17 then. The restricted sample consists of only those adolescents who in addition were found to be enrolled in the second-last grade of middle school in 1999. Treatment effects are estimated for three outcome variables: whether the adolescent did not continue with high school in school year 2000/01 for money-related reasons, whether the adolescent did not continue with high school in school year 2000/01 for reasons other than money, and log monthly per capita expenditure in the adolescent's household.

Point estimates are obtained via the respective orthogonal moment conditions (see appendix section 2.A.2). Standard errors are obtained via wild cluster bootstrap. Results are based on 10-fold cross-fitting, using 100 sample splits. Estimates from the sample splits are combined using the median method, based on equations (2.A.6) and (2.A.7). Only results from combining the best-fitting machine learning methods for each nuisance function are reported here. Estimates of ATE and ATN for all the machine learning methods are reported in section 2.A.7 of the appendix.

### 2.5.3 Coefficient Stability

The estimation of treatment effects using the DML method takes into account a large number of observed characteristics, as well as unobserved characteristics that are correlated with any combination of the observed ones. Nonetheless, the identification is arguably not impervious to any unobserved characteristics. (Due to the lack of an ideal experiment that could be emulated, this holds for any estimation strategy for conditional aftereffects.) Therefore, it seems worth assessing in how far selection bias may remain an issue. One approach to doing so is offered by Oster (2017), who developed an estimator for the omitted variable bias. The idea is to compare the main estimator of interest to the estimate of a simple regression of the outcome on the treatment variable. *Ceteris paribus*, the closer the former is to the latter, the less the observed covariates seem to matter, and the less bias is expected from omitting unobserved ones. Similarly, the better the model with observed covariates explains the outcome and the worse the model without covariates does, the less scope for omitted variable bias is left. In section 2.A.7 of the appendix, I compute a version of Oster’s estimator. It suggests that for all results with significant treatment effects, an inclusion of unobserved factors in the outcome model would not change the ATE estimator enough to cancel out the treatment effect—even if the influence of those factors was up to five times that of the observable factors. Moreover, adding observable characteristics to the simple regression moves the ATE estimate further away from zero. So if the correlations of unobserved portion and observed portion with the treatment variable have the same sign, if anything, one would expect the true effects to be larger in absolute terms than the estimates. In summary, unobserved factors are unlikely to invalidate the findings.

## 2.6 Conclusion

The positive effects of CCT programs like PROGRESA on school enrollment have been demonstrated in numerous studies. However, surprisingly, their aftereffects have not been explored so far. With this chapter, I try to fill this gap by estimating PROGRESA’s impact on the probability to continue school after program payments stop. The main finding is that for the eligible students, the program has large and significant negative aftereffects. There are a number of possible explanations. Financial incentives may crowd out the social norm of sending children to school or reduce the intrinsic motivation to attend school regularly once they stop being in place. Establishing program payments and then reducing them to zero again may convey the false signal that education is not worth it at the later levels. Payments may shift parents’ income reference point such that the

sudden drop needs to be compensated by the child's wage income. And a change in the composition of students induced by the program may lead to negative classroom peer effects. Though conclusive evidence in favor of one over the other explanations is lacking, it seems that loss aversion and classroom peer effects are not much supported by the data, leaving motivation crowding and anchoring as the remaining candidates.

The chapter also looks at possible spillovers to the students who were not eligible to the program but lived in treatment villages. Curiously, it seems that—if anything—these students are more likely to finish high school as a result of their peers getting paid. This could be explained by a heightened desire of the non-poor students to separate themselves from the poor through more education, or to prove their ability despite not getting paid.

Of course, the findings of this chapter are confined to the relatively short time the program had been in effect by the year 2000. For younger cohorts who start middle school with the program already in place, one may expect positive unconditional treatment effects, via the intermediate positive effect the program has on middle school enrollment. On the other hand, the conditional treatment effects might be even more extreme due to a longer exposure to the program.

The main result is remarkable, as it constitutes a textbook case of unintended consequences. It encourages to look further into how motivation and social norms change through financial incentives. It raises the question of whether a potential motivation crowding effect carries on to higher education, vocational training, or the labor market, and whether it shows not only in participation but also performance. Moreover, the finding should be considered in the program design of future CCT programs. Even in cases where coverage on all school levels is not feasible due to budgetary constraints, there may be ways to counter the adverse program effect. This could, for instance, be done by systematically informing students and parents about the marginal rate of return to continued education. Another way may be to let go of the conditionality of payments altogether, particularly to counter possible crowding-out effects and anchoring. Further research that explores these channels may help to mitigate the negative side effects of CCT programs.

## 2.A Appendix

### 2.A.1 List of Pre-treatment Characteristics

Table 2.A.1: Pre-treatment characteristics

Variable description	Type
<i>Student and household characteristics</i>	
Student is female	binary
Age of student in 1997	continuous
Degree of poverty index in 1997 (by 1997 criteria)	continuous
Degree of poverty index in 1997 (by 2003 criteria)	continuous
Very poor, poor, marginally non-poor, or clearly non-poor in 1997 (by 1997 criteria)	categorical
Household size	count
Number of household members below age 15	count
Father lives in the household	binary
Mother lives in the household	binary
Father is literate	binary
Mother is literate	binary
Father went to school	binary
Mother went to school	binary
Father finished at least primary school	binary
Mother finished at least primary school	binary
Father finished at least middle school	binary
Mother finished at least middle school	binary
Student attended school in October 1997	binary
Student attended school in March 1998	binary
<i>Parents' assessments and opinions</i>	
Parents say student is good at school in 1998	binary
Parents say student is able to finish middle school	binary
Parents say student is able to continue after middle school	binary
Parents say student can finish high school	binary
Parents say student is able to finish university	binary
Desired level of schooling for girls is at least middle school	binary
Desired level of schooling for girls is more than middle school	binary
Desired level of schooling for girls is at least high school	binary



Table 2.A.1: Pre-treatment characteristics

Variable description	Type
Desired level of schooling for girls is university	binary
Desired level of schooling for boys is at least middle school	binary
Desired level of schooling for boys is more than middle school	binary
Desired level of schooling for boys is at least high school	binary
Desired level of schooling for boys is university	binary
Children eat breakfast before school	binary
Reason why children don't eat breakfast before school	categorical
Parent talked to teacher this year	binary
Reason for talk with teacher	categorical
Parent participates in parent / guardian association of school	binary
Parent participates in school work	binary
In school, there are problems with lack of discipline	binary
In school, there are problems with lack of interest of the teachers	binary
In school, there are problems with poor communication between teachers and parents	binary
In school, there are problems with poor teacher attendance	binary
The teacher is usually prepared	binary
The teacher is usually fulfilled	binary
The teacher is usually on time	binary
The teacher is usually patient with the children	binary
Age from which girls can help younger siblings	continuous
Age from which boys can help younger siblings	continuous
Age from which girls can help with work	continuous
Age from which boys can help with work	continuous
Age from which girls can work to earn money	continuous
Age from which boys can work to earn money	continuous
<i>Household expenditures</i>	
Weekly expenditures for public transport to school	continuous
Weekly expenditures for public transport for other trips	continuous
Weekly expenditures for cigarettes and tobacco	continuous
Weekly expenditures for alcoholic beverages	continuous
Weekly expenditures for nonalcoholic beverages	continuous
Monthly expenditures for hygiene items	continuous

Table 2.A.1: Pre-treatment characteristics

Variable description	Type
Monthly expenditures for medicine	continuous
Monthly expenditures for medical consultations	continuous
Biannual expenditures for household articles	continuous
Biannual expenditures for toys	continuous
Biannual expenditures for girls' clothes	continuous
Biannual expenditures for boys' clothes	continuous
Biannual expenditures for women's clothes	continuous
Biannual expenditures for men's clothes	continuous
Biannual expenditures for girls' shoes	continuous
Biannual expenditures for boys' shoes	continuous
Biannual expenditures for women's shoes	continuous
Biannual expenditures for men's shoes	continuous
Biannual expenditures for school supplies	continuous
Biannual expenditures for school contributions	continuous
If family had more money, they would spend it on food	rank
If family had more money, they would spend it on housing repairs	rank
If family had more money, they would spend it on clothing or shoes	rank
If family had more money, they would spend it on debt settlement	rank
If family had more money, they would spend it on animals	rank
If family had more money, they would spend it on seeds or plants	rank
If family had more money, they would spend it on work tools	rank
If family had more money, they would spend it on medicine	rank
If family had more money, they would spend it on school supplies	rank
If family had more money, they would save it	rank
<i>Location characteristics</i>	
Marginality index	continuous
Degree of marginality very high (1) or high (0) in 1997	binary
Village is indigenous	binary
Village has a municipal delegate	binary
Village has a municipal subdelegate	binary
Village has a commissioner of agricultural land	binary
Village has a commissioner of communal goods	binary
Village has a municipal development committee	binary

Table 2.A.1: Pre-treatment characteristics

Variable description	Type
Village has a health committee	binary
Village has a education committee	binary
Village has a agricultural committee	binary
Village has a DICONSA store officer	binary
Village has a production cooperative	binary
Village has religious organizations	binary
Village has political organizations	binary
Village has a school parent association	binary
Village has community assemblies	binary
Village has NGOs	binary
Village has a communal work system (tequio)	binary
Source of water	categorical
Type of garbage disposal	categorical
Electricity available everywhere	binary
Electricity at least partly available	binary
Public drainage at least partly available	binary
Public phone available	binary
Number of preschools	count
Number of primary schools	count
Number of distance middle schools	count
Most important sector in this village	categorical
Second most important sector in this village	categorical
Third most important sector in this village	categorical
Child labor takes place in this village	binary
Average daily salary paid to children	continuous
Number of inhabitants	count
Number of poor inhabitants	count
Number of primary school graduates between 11 and 14	count
Share among inhabitants of primary school graduates between 11 and 14	share
Number of poor primary school graduates between 11 and 14	count
Share among poor inhabitants of primary school graduates between 11 and 14	share

Table 2.A.1: Pre-treatment characteristics

Variable description	Type
Number of primary school graduates between 11 and 14 enrolled in 1997	count
Share among inhabitants of primary school graduates between 11 and 14 enrolled in 1997	share
Number of poor primary school graduates between 11 and 14 enrolled in 1997	count
Share among poor inhabitants of prim. school graduates between 11 and 14 enrolled in 1997	share
Number of inhabitants who completed at least primary school	count
Share among inhabitants who completed at least primary school	share
Share among inhabitants between 15 and 20 who completed at least primary school	share
Share among inhabitants between 21 and 30 who completed at least primary school	share
Number of poor inhabitants who completed at least primary school	count
Share among poor inhabitants who completed at least primary school	share
Share among poor inhabitants between 15 and 20 who completed at least primary school	share
Share among poor inhabitants between 21 and 30 who completed at least primary school	share
Number of inhabitants who completed at least secondary school	count
Share among inhabitants who completed at least secondary school	share
Share among inhabitants between 15 and 20 who completed at least secondary school	share
Share among inhabitants between 21 and 30 who completed at least secondary school	share
Number of poor inhabitants who completed at least secondary school	count
Share among poor inhabitants who completed at least secondary school	share
Share among poor inhabitants between 15 and 20 who completed at least secondary school	share
Share among poor inhabitants between 21 and 30 who completed at least secondary school	share
Number of inhabitants who completed at least high school	count

Table 2.A.1: Pre-treatment characteristics

Variable description	Type
Share among inhabitants who completed at least high school	share
Share among inhabitants between 15 and 20 who completed at least high school	share
Share among inhabitants between 21 and 30 who completed at least high school	share
Number of poor inhabitants who completed at least high school	count
Share among poor inhabitants who completed at least high school	share
Share among poor inhabitants between 15 and 20 who completed at least high school	share
Share among poor inhabitants between 21 and 30 who completed at least high school	share
Travel time (minutes) to nearest private middle school	continuous
Travel time (minutes) to nearest public middle school	continuous
Travel time (minutes) to nearest distance middle school	continuous
Travel time (minutes) to nearest middle school	continuous
Travel time (minutes) to nearest private high school	continuous
Travel time (minutes) to nearest public high school	continuous
Travel time (minutes) to nearest high school	continuous
Travel time (minutes) to nearest national college of technical professional education	continuous
Travel time (minutes) to nearest agricultural technological center	continuous
Travel time (minutes) to nearest industrial technology and services center	continuous
Travel time (minutes) to nearest agricultural college	continuous
Travel time (minutes) to nearest industrial and services college	continuous

## 2.A.2 Score Functions

For the model without attrition, Chernozhukov et al. (2018) show that under a number of regularity conditions—particularly concerning the speed at which the nuisance functions  $\eta$  converge to their true values  $\eta_0$ —and using Neyman-orthogonal moment conditions as well as cross-fitting, their estimators of the ATE and the ATT (average treatment effect of the treated) are  $\sqrt{N}$ -consistent and asymptotically normal. The authors state that a

crude requirement for the nuisance functions is that they converge at rate  $o(N^{-1/4})$ . This rate is shown to be achievable for a variety of data generating processes in conjunction with specific machine learning methods. Given the possibility to aggregate or choose the best out of multiple machine learning methods, this guarantees estimability for a wide range of problems.

A Neyman-orthogonal score function for the ATE is

$$\begin{aligned} \psi_{\theta}(W; \theta, \eta) = & g(1, X) - g(0, X) + \frac{D(Y - g(1, X))}{m(X)} \\ & - \frac{(1 - D)(Y - g(0, X))}{1 - m(X)} - \theta, \end{aligned} \quad (2.A.1)$$

with data  $W = (Y, D, X)$  and nuisance functions  $\eta(X) = (g(1, X), g(0, X), m(X))$ . The true value of  $\eta$  is  $\eta_0(X) = (g_0(1, X), g_0(0, X), m_0(X))$ . A Neyman-orthogonal score function for the ATN is

$$\begin{aligned} \psi_{\gamma}(W; \gamma, \eta) = & \frac{D(1 - m(X))(Y - g(1, X))}{m(X)(1 - p_D)} - \frac{(1 - D)(Y - g(0, X))}{1 - p_D} \\ & + \frac{(1 - D)(g(1, X) - g(0, X))}{1 - p_D} - \gamma \frac{1 - D}{1 - p_D}, \end{aligned} \quad (2.A.2)$$

with nuisance functions  $\eta(X) = (g(1, X), g(0, X), m(X), p_D)$ . Here, the true value of  $\eta$  is  $\eta_0(X) = (g_0(1, X), g_0(0, X), m_0(X), E[D])$ .

For the model with attrition, the moment conditions need to be adapted to the fact that not all outcomes are observed. This entails that treated observations with observed outcomes are weighted by the inverse conditional probability of being observed and treated. Non-treated observations with observed outcomes are weighted by the inverse conditional probability of being observed and non-treated.

Let  $q_0(D, X) := Dr_0(1, X)m_0(X) + (1 - D)r_0(0, X)(1 - m_0(X))$ . Then, the corresponding score function for the ATE is

$$\begin{aligned} \varphi_{\theta}(W; \theta, \eta) = & g(1, X) - g(0, X) + \frac{RD(Y - g(1, X))}{q(1, X)} \\ & - \frac{R(1 - D)(Y - g(0, X))}{q(0, X)} - \theta, \end{aligned} \quad (2.A.3)$$

with data  $W = (Y, D, R, X)$  and nuisance functions  $\eta(X) = (g(1, X), g(0, X), q(1, X), q(0, X))$  whose true value is  $\eta_0(X) = (g_0(1, X), g_0(0, X), r_0(1, X)m_0(X), r_0(0, X)(1 - m_0(X)))$ .

For the ATN, the corresponding score function is

$$\begin{aligned} \varphi_\gamma(W; \gamma, \eta) = & \frac{RD(Y - g(1, X))(1 - m(X))}{q(1, X)(1 - p_D)} - \frac{R(1 - D)(Y - g(0, X))}{r(0, X)(1 - p_D)} \\ & + \frac{(1 - D)(g(1, X) - g(0, X))}{1 - p_D} - \gamma \frac{1 - D}{1 - p_D}, \end{aligned} \quad (2.A.4)$$

with nuisance functions  $\eta(X) = (g(1, X), g(0, X), m(X), r(0, X), q(1, X), p_D)$ , having true value  $\eta_0(X) = (g_0(1, X), g_0(0, X), m_0(X), r_0(0, X), r_0(1, X), m_0(X), E[D])$ .

### 2.A.3 Repeated Cross-fitting

The cross-fitting procedure works the same way for all score functions; I use  $\psi_\theta$  as the example here. For a fixed integer  $K$ , the sample is randomly split into folds  $I_1, \dots, I_K$  of roughly equal size with no overlapping locations. For each  $k \in \{1, \dots, K\}$ , the nuisance functions are estimated using only the observations outside of  $I_k$ . The resulting functional estimates are then used to predict  $\eta_0(X)$  in fold  $I_k$ . The predictions over all folds are in turn used to obtain the point estimate of  $\theta$ , through the equation

$$E \left[ \psi_\theta(W; \hat{\theta}, \hat{\eta}) \right] = 0. \quad (2.A.5)$$

The sample-splitting procedure itself also introduces additional uncertainty. Therefore, the above procedure is repeated a number of times  $B$  with different random splits. The final estimator is then put together via the median method suggested by Chernozhukov et al. (2018). The final point estimate is the median of estimates for each split,

$$\hat{\theta}^{\text{median}} = \text{median} \left\{ \hat{\theta}_b \right\}_{b=1}^B. \quad (2.A.6)$$

The final variance estimator takes into account the variation introduced by sample splitting:

$$\hat{\sigma}^{2, \text{median}} = \text{median} \left\{ \hat{\sigma}_b^2 + \left( \hat{\theta}_b - \hat{\theta}^{\text{median}} \right)^2 \right\}_{b=1}^B. \quad (2.A.7)$$

In this chapter, I obtain  $\hat{\sigma}_b^2$  via a wild cluster bootstrap over the values of  $\psi_\theta(W; \hat{\theta}_b, \hat{\eta}_b)$ .

### 2.A.4 Inference for the Model with Attrition

The following theorem parallels Theorem 5.1 in Chernozhukov et al. (2018), stating that the DML estimators of the ATE and the ATN for the model with attrition are approximately unbiased and asymptotically normal.

Expectation and probability operators as well as norms are always with respect to a probability measure  $P$  of the data  $W = (Y, D, R, X)$ . I use  $\|\cdot\|_q$  to denote the  $L^q(P)$  norm, and for nuisance functions  $\eta = (\ell_1, \dots, \ell_l)$ , denote  $\|\eta\|_q := \max_{1 \leq j \leq l} \|\ell_j\|_q$ . Let  $(\delta_n)_{n=1}^\infty$  and  $(\Delta_n)_{n=1}^\infty$  be sequences of positive constants approaching 0, and let  $\varepsilon, c, C, C'$ , and  $q$  be positive constants, with  $q > 2$ .

**THEOREM.** Assume that the following conditions hold: (a) equations (2.3.10) – (2.3.12) hold; (b)  $\|Y\|_q \leq C$ ; (c)  $\Pr(\varepsilon \leq q_0(D, X) \leq 1 - \varepsilon) = 1$ ; (d)  $\Pr(\varepsilon \leq r_0(D, X)) = 1$ ; (e)  $\|RU\|_2 \geq c$ ; (f)  $\|E[U^2|X]\|_\infty \leq C$ ; (g) for subset  $I$  of  $[N]$  of size  $n$ ,  $\eta = \eta((W_i)_{i \in I}) \in \mathcal{T}_N$  with  $P$ -probability no less than  $1 - \Delta_N$ , where the realization set  $\mathcal{T}_N$  is a shrinking neighborhood of  $\eta_0$  containing all the nuisance parameter estimates  $\eta$  that obey the following conditions:  $\|\eta - \eta_0\|_q \leq C$ ,  $\|\eta - \eta_0\|_2 \leq \delta_N$ ,  $\max\{\|m - \frac{1}{2}\|_\infty, \|q - \frac{1}{2}\|_\infty\} \leq \frac{1}{2} - \varepsilon$ ,  $\|r\|_\infty \geq \varepsilon$ , and  $\|g - g_0\|_2 \times (\|m - m_0\|_2 + \|r - r_0\|_2 + \|q - q_0\|_2) \leq \delta_N N^{-1/2}$ . Then, the DML estimators for the ATE and ATN constructed above,  $\hat{\theta}_0$  and  $\hat{\gamma}_0$ , obey  $\sqrt{N}(\hat{\theta}_0 - \theta_0) \rightsquigarrow \mathcal{N}(0, \sigma_\theta^2)$  with  $\sigma_\theta^2 = E[\varphi_\theta^2(W; \theta_0, \eta_0)]$  as well as  $\sqrt{N}(\hat{\gamma}_0 - \gamma_0) \rightsquigarrow \mathcal{N}(0, \sigma_\gamma^2)$  with  $\sigma_\gamma^2 = E[\varphi_\gamma^2(W; \gamma_0, \eta_0)]$ .

**PROOF:** The proof follows along the same lines as the one given in Chernozhukov et al. (2018, pp. 65–68) for the model without attrition. I confine myself to showing it for the ATN, omitting the ATE. It suffices to show that the score function  $\varphi_\gamma$  fulfills the conditions given in Assumptions 3.1 and 3.2 in Chernozhukov et al. These conditions are Neyman-orthogonality, identifiability, non-degenerate variance, and some conditions on the quality of the nuisance parameter estimates. I go through these conditions separately. Once they are established, the above theorem follows from Theorem 3.1 in Chernozhukov et al.

**Neyman-orthogonality** Neyman-orthogonality of the score function means that at the true parameter values  $\gamma_0$  and  $\eta_0$ ,  $\varphi_\gamma$  obeys the moment conditions  $E[\varphi_\gamma(W; \gamma_0, \eta_0)] = 0$ , and that for nuisance parameter estimate  $\eta \in \mathcal{T}_N$ , the Gateaux derivative in the direction  $\eta - \eta_0$  is zero. The first part is true by definition of  $\gamma_0$ . As for the second part, it holds that



$$\begin{aligned}
\frac{\partial}{\partial \eta} \mathbb{E} [\varphi_\gamma (W; \gamma_0, \eta_0)] [\eta - \eta_0] &= \frac{\partial}{\partial t} \mathbb{E} [\varphi_\gamma (W; \gamma_0, \eta_0 + t(\eta - \eta_0))] |_{t=0} \\
&= \mathbb{E} \left[ \frac{DR(g_0(1, X) - g(1, X))(1 - m_0(X))}{q_0(1, X)(1 - \mathbb{E}[D])} \right] \\
&\quad + \mathbb{E} \left[ \frac{DR(Y - g_0(1, X))(1 - m_0(X))(q_0(1, X) - q(1, X))}{q_0(1, X)^2(1 - \mathbb{E}[D])} \right] \\
&\quad + \mathbb{E} \left[ \frac{DR(Y - g_0(1, X))(m_0(X) - m(X))}{q_0(1, X)(1 - \mathbb{E}[D])} \right] \\
&\quad - \mathbb{E} \left[ \frac{DR(Y - g_0(1, X))(1 - m_0(X))(\mathbb{E}[D] - p_D)}{q_0(1, X)(1 - \mathbb{E}[D])^2} \right] \\
&\quad - \mathbb{E} \left[ \frac{(1 - D)R(g_0(0, X) - g(0, X))}{r_0(0, X)(1 - \mathbb{E}[D])} \right] \\
&\quad - \mathbb{E} \left[ \frac{(1 - D)R(Y - g_0(0, X))(r_0(0, X) - r(0, X))}{r_0(0, X)^2(1 - \mathbb{E}[D])} \right] \\
&\quad + \mathbb{E} \left[ \frac{(1 - D)R(Y - g_0(0, X))(\mathbb{E}[D] - p_D)}{r_0(0, X)(1 - \mathbb{E}[D])^2} \right] \\
&\quad - \mathbb{E} \left[ \frac{(1 - D)(g_0(1, X) - g(1, X) - g_0(0, X) + g(0, X))}{1 - \mathbb{E}[D]} \right] \\
&\quad - \mathbb{E} \left[ \frac{(1 - D)(g_0(1, X) - g_0(0, X))(\mathbb{E}[D] - p_D)}{(1 - \mathbb{E}[D])^2} \right] \\
&\quad + \mathbb{E} \left[ \frac{(1 - D)(\mathbb{E}[D] - p_D)\gamma_0}{(1 - \mathbb{E}[D])^2} \right] \\
&= \mathbb{E} \left[ \frac{(g_0(1, X) - g(1, X))(1 - m_0(X))}{1 - \mathbb{E}[D]} \right] + 0 + 0 - 0 \\
&\quad - \mathbb{E} \left[ \frac{(g_0(0, X) - g(0, X))(1 - m_0(X))}{1 - \mathbb{E}[D]} \right] - 0 + 0 \\
&\quad - \mathbb{E} \left[ \frac{(1 - m_0(X))(g_0(1, X) - g(1, X) - g_0(0, X) + g(0, X))}{1 - \mathbb{E}[D]} \right] \\
&\quad - \frac{\mathbb{E}[D] - p_D}{(1 - \mathbb{E}[D])^2} \mathbb{E}[(1 - D)(g_0(1, X) - g_0(0, X) - \gamma_0)] \\
&= 0
\end{aligned}$$

by the law of iterated expectations, since  $\mathbb{E}[Y - g_0(1, X) | D = 1] = 0$ ,

$$\mathbb{E}[Y - g_0(0, X) | D = 0] = 0, \quad \mathbb{E} \left[ \frac{DR}{q_0(1, X)} | X \right] = 1,$$

$$\mathbb{E} \left[ \frac{(1-D)R}{r_0(0, X)} | X \right] = \mathbb{E}[(1 - D) | X] = 1 - m_0(X),$$

$$\text{and } \mathbb{E}[g_0(1, X) - g_0(0, X) - \gamma_0 | D = 0] = 0.$$

**Identifiability** Next, for identifiability, note that  $\varphi_\gamma$  is linear in  $\gamma$  with coefficient  $\varphi_\gamma^a(W; \eta) = -\frac{1-D}{1-p_D}$ . In order for  $\gamma$  to be identified,  $\varphi_\gamma^a(W; \eta)$  needs to be a nonzero real number on expectation. This is true here, since  $E\left[-\frac{1-D}{1-p_D}\right] = -1$ .

**Non-degenerate Variance** Another condition is that the score variance is non-degenerate, i.e.,  $E[\varphi_\gamma^2(W; \gamma_0, \eta_0)]$  is bounded away from zero:

$$\begin{aligned}
E[\varphi_\gamma^2(W; \gamma_0, \eta_0)] &= E\left[\left(\frac{RD(Y - g_0(1, X))(1 - m_0(X))}{q_0(1, X)(1 - E[D])} - \frac{R(1 - D)(Y - g_0(0, X))}{r_0(0, X)(1 - E[D])}\right)^2\right] \\
&\quad + E\left[\left(\frac{(1 - D)(g_0(1, X) - g_0(0, X))}{1 - E[D]} - \gamma \frac{1 - D}{1 - E[D]}\right)^2\right] \\
&\geq E\left[\left(\frac{RD(Y - g_0(1, X))(1 - m_0(X))}{q_0(1, X)(1 - E[D])} - \frac{R(1 - D)(Y - g_0(0, X))}{r_0(0, X)(1 - E[D])}\right)^2\right] \\
&= E\left[\frac{RD(Y - g_0(1, X))^2(1 - m_0(X))^2}{q_0(1, X)^2(1 - E[D])^2} + \frac{R(1 - D)(Y - g_0(0, X))^2}{r_0(0, X)^2(1 - E[D])^2}\right] \\
&= E\left[\frac{RD(Y - g_0(1, X))^2(1 - m_0(X))^2}{q_0(1, X)^2(1 - E[D])^2}\right. \\
&\quad \left.+ \frac{R(1 - D)(Y - g_0(0, X))^2(1 - m_0(X))^2}{q_0(0, X)^2(1 - E[D])^2}\right] \\
&= E\left[\frac{R(1 - m_0(X))^2}{(1 - E[D])^2} \left(\frac{D(Y - g_0(1, X))^2}{q_0(1, X)^2} + \frac{(1 - D)(Y - g_0(0, X))^2}{q_0(0, X)^2}\right)\right] \\
&\geq E\left[\frac{R(1 - m_0(X))^2}{(1 - E[D])^2} (D(Y - g_0(1, X))^2 + (1 - D)(Y - g_0(0, X))^2)\right] \\
&\geq \frac{\varepsilon^2}{(1 - E[D])^2} E[RU^2] \geq \frac{\varepsilon^2 c}{(1 - E[D])^2} > 0.
\end{aligned}$$

**Existence of Moments** The last set of conditions concerns the quality of the nuisance parameter estimators  $\eta \in \mathcal{T}_N$ . The first condition is the existence of all moments of  $\varphi_\gamma(W; \gamma_0, \eta)$  and  $\varphi_\gamma^a(W; \eta)$ , meaning that  $\sup_{\eta \in \mathcal{T}_N} \|\varphi_\gamma(W; \gamma_0, \eta)\|_q$  and  $\sup_{\eta \in \mathcal{T}_N} \|\varphi_\gamma^a(W; \eta)\|_q$  are bounded from above. It holds that

$$\begin{aligned}
E[|\varphi_\gamma(W; \gamma_0, \eta)|^q]^{\frac{1}{q}} &= \|\varphi_\gamma(W; \gamma_0, \eta)\|_q \\
&= \left\| \frac{RD(Y - g(1, X))(1 - m(X))}{q(1, X)(1 - p_D)} - \frac{R(1 - D)(Y - g(0, X))}{r(0, X)(1 - p_D)} \right\|_q
\end{aligned}$$

$$\begin{aligned}
& + \left\| \frac{(1-D)(g(1, X) - g(0, X))}{1-p_D} - \gamma \frac{1-D}{1-p_D} \right\|_q \\
& \leq \left\| \frac{RD(Y - g(1, X))(1-m(X))}{q(1, X)(1-p_D)} \right\|_q + \left\| \frac{R(1-D)(Y - g(0, X))}{r(0, X)(1-p_D)} \right\|_q \\
& \quad + \left\| \frac{(1-D)(g(1, X) - g(0, X))}{1-p_D} \right\|_q + \left\| \gamma \frac{1-D}{1-p_D} \right\|_q \\
& \leq \frac{1}{1-p_D} \left( \frac{1-\varepsilon}{\varepsilon} (\|Y\|_q + \|g(1, X)\|_q) \right. \\
& \quad \left. + \frac{1}{\varepsilon} (\|Y\|_q + \|g(0, X)\|_q) + \|g(1, X) - g(0, X)\|_q + |\gamma_0| \right) \\
& \leq \frac{1}{1-p_D} \frac{1}{\varepsilon} \left( \|g(1, X) - g_0(1, X)\|_q + \|g_0(1, X)\|_q \right) \\
& \quad + \frac{1}{1-p_D} \frac{1+\varepsilon}{\varepsilon} \left( \|g(0, X) - g_0(0, X)\|_q + \|g_0(0, X)\|_q \right) \\
& \quad + \frac{1}{1-p_D} \left( \frac{2-\varepsilon}{\varepsilon} \|Y\|_q + |\gamma_0| \right),
\end{aligned}$$

$\|Y\|_q$  and  $|\gamma_0|$  are bounded by assumption, and Chernozhukov et al. (2018, pp. 66–67) show that  $\|g_0(0, X)\|_q$ ,  $\|g_0(1, X)\|_q$ ,  $\|g(0, X) - g_0(0, X)\|_q$ , and  $\|g(1, X) - g_0(1, X)\|_q$  are bounded by  $C/\varepsilon^{\frac{1}{q}}$ .

Furthermore, it holds that  $E[|\varphi_\gamma^a(W; \eta)|^q]^{\frac{1}{q}} = \frac{1}{1-p_D} E[1-D]^{\frac{1}{q}} < \frac{1}{1-p_D}$ .

**Statistical Rates** The remaining conditions characterize the statistical rates with which the realization set  $\mathcal{T}_N$  shrinks:

$$\sup_{\eta \in \mathcal{T}_N} |E[\varphi_\gamma^a(W; \eta)] - E[\varphi_\gamma^a(W; \eta_0)]| \leq C' \delta_N, \quad (2.A.8)$$

$$\sup_{\eta \in \mathcal{T}_N} \|\varphi_\gamma(W; \gamma_0, \eta) - \varphi_\gamma(W; \gamma_0, \eta_0)\|_2 \leq C' \delta_N, \quad (2.A.9)$$

$$\sup_{t \in (0,1), \eta \in \mathcal{T}_N} \left| \frac{\partial^2}{\partial t^2} E[\varphi_\gamma(W; \gamma_0, \eta_0 + t(\eta - \eta_0))] \right| \leq C' \delta_N / \sqrt{N}. \quad (2.A.10)$$

For  $\eta \in \mathcal{T}_N$  it holds that  $|E[\varphi_\gamma^a(W; \eta)] - E[\varphi_\gamma^a(W; \eta_0)]| = \left| \frac{1-E[D]}{1-p_D} - 1 \right| \leq \frac{\delta_N}{1-p_D} \leq C' \delta_N$  for  $C' \geq \frac{1}{1-p_D}$ , which gives (2.A.8). To show (2.A.9), note that by the triangle inequality,

$$\|\varphi_\gamma(W; \gamma_0, \eta) - \varphi_\gamma(W; \gamma_0, \eta_0)\|_2 \leq \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4,$$

with

$$\begin{aligned}\mathcal{I}_1 &:= \left\| \frac{RD(Y-g(1,X))(1-m(X))}{q(1,X)(1-p_D)} - \frac{RD(Y-g_0(1,X))(1-m_0(X))}{q_0(1,X)(1-E[D])} \right\|_2, \\ \mathcal{I}_2 &:= \left\| \frac{R(1-D)(Y-g(0,X))}{r(0,X)(1-p_D)} - \frac{R(1-D)(Y-g_0(0,X))}{r_0(0,X)(1-E[D])} \right\|_2, \\ \mathcal{I}_3 &:= \left\| \frac{(1-D)(g(1,X)-g(0,X))}{1-p_D} - \frac{(1-D)(g_0(1,X)-g_0(0,X))}{1-E[D]} \right\|_2, \\ \mathcal{I}_4 &:= \left\| \gamma \frac{1-D}{1-p_D} - \gamma \frac{1-D}{1-E[D]} \right\|_2.\end{aligned}$$

For  $\mathcal{I}_1$ , it holds that

$$\begin{aligned}\mathcal{I}_1 &\leq \varepsilon^{-2} (1-p_{\max})^{-2} \|RD(Y-g(1,X))(1-m(X))q_0(1,X)(1-E[D]) \\ &\quad - RD(Y-g_0(1,X))(1-m_0(X))q(1,X)(1-p_D)\|_2 \\ &\leq \varepsilon^{-2} (1-p_{\max})^{-2} \|(g_0(1,X)+U-g(1,X))(1-m(X))q_0(1,X)(1-E[D]) \\ &\quad - U(1-m_0(X))q(1,X)(1-p_D)\|_2 \\ &\leq \varepsilon^{-2} (1-p_{\max})^{-2} (\|(g(1,X)-g_0(1,X))(1-m(X))q_0(1,X)(1-E[D])\|_2 \\ &\quad + \|U((1-m(X))q_0(1,X)(1-E[D])-(1-m_0(X))q(1,X)(1-p_D))\|_2) \\ &\leq \varepsilon^{-2} (1-p_{\max})^{-2} \left( \|(g(1,X)-g_0(1,X))\|_2 \right. \\ &\quad + \sqrt{C} \|(1-m_0(X)+m_0(X)-m(X))q_0(1,X)(1-E[D]) \\ &\quad \left. - (1-m_0(X))(q_0(1,X)+q(1,X)-q_0(1,X))(1-E[D]-p_D+E[D])\|_2 \right) \\ &\leq \varepsilon^{-2} (1-p_{\max})^{-2} \left( \varepsilon^{-\frac{1}{2}}\delta_N + \sqrt{C} \|(m_0(X)-m(X))q_0(1,X)(1-E[D]) \right. \\ &\quad + (1-m_0(X))(q_0(1,X))(p_D-E[D]) - (1-m_0(X))(q(1,X)-q_0(1,X))(1-E[D]) \\ &\quad \left. + (1-m_0(X))(q(1,X)-q_0(1,X))(p_D-E[D])\|_2 \right) \\ &\leq \varepsilon^{-2} (1-p_{\max})^{-2} \left( \varepsilon^{-\frac{1}{2}}\delta_N + \sqrt{C} (\|m_0(X)-m(X)\|_2 + \|p_D-E[D]\|_2 \right. \\ &\quad \left. + \|q(1,X)-q_0(1,X)\|_2 + \|(q(1,X)-q_0(1,X))(p_D-E[D])\|_2 \right) \\ &\leq \varepsilon^{-2} (1-p_D)^{-2} \left( \varepsilon^{-\frac{1}{2}} + 4\sqrt{C} \right) \delta_N,\end{aligned}$$

with  $p_{\max} := \max\{p_D, E[D]\}$ . Similarly,

$$\begin{aligned}\mathcal{I}_2 &\leq \varepsilon^{-2} (1-p_D)^{-2} \|R(1-D)(Y-g(0,X))r_0(0,X)(1-E[D]) \\ &\quad - R(1-D)(Y-g_0(0,X))r(0,X)(1-p_D)\|_2 \\ &\leq \varepsilon^{-2} (1-p_D)^{-2} \|(g_0(0,X)+U-g(0,X))r_0(0,X)(1-E[D]) - Ur(0,X)(1-p_D)\|_2 \\ &\leq \varepsilon^{-2} (1-p_D)^{-2} \|(g_0(0,X)-g(0,X))r_0(0,X)(1-E[D]) \\ &\quad + U(r_0(0,X)(1-E[D]) - r(0,X)(1-p_D))\|_2 \\ &\leq \varepsilon^{-2} (1-p_D)^{-2} \left( \|(g_0(0,X)-g(0,X))\|_2 + \sqrt{C} \|r_0(0,X)(1-E[D]) - r(0,X)(1-p_D)\|_2 \right)\end{aligned}$$

$$\begin{aligned}
&\leq \varepsilon^{-2} (1 - p_D)^{-2} \left( \| (g_0(0, X) - g(0, X)) \|_2 + \sqrt{C} \| r_0(0, X) (1 - \mathbb{E}[D]) \right. \\
&\quad \left. - r(0, X) (1 - \mathbb{E}[D]) - r(0, X) (\mathbb{E}[D] - p_D) \|_2 \right) \\
&\leq \varepsilon^{-2} (1 - p_D)^{-2} \left( \varepsilon^{-\frac{1}{2}} \delta_N + \sqrt{C} (\| r_0(0, X) - r(0, X) \|_2 + \| \mathbb{E}[D] - p_D \|_2) \right) \\
&\leq \varepsilon^{-2} (1 - p_D)^{-2} \left( \varepsilon^{-\frac{1}{2}} + 2\sqrt{C} \right) \delta_N,
\end{aligned}$$

$$\begin{aligned}
\mathcal{I}_3 &\leq (1 - p_D)^{-2} \| (1 - D) (g(1, X) - g(0, X)) (1 - \mathbb{E}[D]) \\
&\quad - (1 - D) (g_0(1, X) - g_0(0, X)) (1 - p_D) \|_2 \\
&\leq (1 - p_D)^{-2} \| (g(1, X) - g_0(1, X) - g(0, X) + g_0(0, X)) (1 - \mathbb{E}[D]) \\
&\quad - (g_0(1, X) - g_0(0, X)) (\mathbb{E}[D] - p_D) \|_2 \\
&\leq (1 - p_D)^{-2} (\| g(1, X) - g_0(1, X) \|_2 + \| g(0, X) - g_0(0, X) \|_2 \\
&\quad + \| g_0(1, X) \|_2 + \| g_0(0, X) \|_2) \\
&\leq (1 - p_D)^{-2} 4\varepsilon^{-\frac{1}{2}} \delta_N,
\end{aligned}$$

$$\begin{aligned}
\mathcal{I}_4 &\leq (1 - p_D)^{-2} \| \gamma_0 (1 - D) (p_D - \mathbb{E}[D]) \|_2 \\
&\leq (1 - p_D)^{-2} \gamma_0 \delta_N.
\end{aligned}$$

Thus, for  $C' \geq (1 - p_D)^{-2} \left( 2\varepsilon^{-\frac{5}{2}} + 6\sqrt{C}\varepsilon^{-2} + 4\varepsilon^{-\frac{1}{2}} + \gamma_0 \right)$  we have

$$\| \varphi_\gamma(W; \gamma_0, \eta) - \varphi_\gamma(W; \gamma_0, \eta_0) \|_2 \leq (1 - p_D)^{-2} \left( 2\varepsilon^{-\frac{5}{2}} + 6\sqrt{C}\varepsilon^{-2} + 4\varepsilon^{-\frac{1}{2}} + \gamma_0 \right) \delta_N \leq C' \delta_N.$$

Lastly, for  $t \in (0, 1)$ ,  $\frac{\partial^2}{\partial t^2} \mathbb{E}[\varphi_\gamma(W; \gamma_0, \eta_0 + t(\eta - \eta_0))] = \mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3 + \mathcal{J}_4$  with

$$\begin{aligned}
\mathcal{J}_1 &:= 2\mathbb{E} \left[ DR \left\{ (p_D - \mathbb{E}[D]) (m_0(X) - m(X) + p_D - \mathbb{E}[D]) + (m(X) - m_0(X)) \mathbb{E}[D] + m_0(X) (\mathbb{E}[D] - p_D) \right. \right. \\
&\quad \times (q_0(1, X) + t(q(1, X) - q_0(1, X)))^2 (Y - g_0(1, X) - t(g(1, X) - g_0(1, X))) \\
&\quad + ((Y - g_0(1, X)) (q(1, X) - q_0(1, X)) + (g(1, X) - g_0(1, X)) q_0(1, X)) \\
&\quad \times \left( ((m(X) - m_0(X)) (1 - \mathbb{E}[D]) + \mathbb{E}[D] - p_D - m_0(X) (\mathbb{E}[D] - p_D)) \right. \\
&\quad \times (\mathbb{E}[D] - 1 + t(p_D - \mathbb{E}[D])) ((q_0(1, X) - q(1, X)) t - q_0(1, X)) \\
&\quad \left. \left. + (q(1, X) - q_0(1, X)) (1 - m_0(X) + t(m_0(X) - m(X))) (1 - \mathbb{E}[D] + t(\mathbb{E}[D] - p_D))^2 \right) \right\} \\
&\quad \left. / \left( (1 - \mathbb{E}[D] + t(\mathbb{E}[D] - p_D))^3 (q_0(1, X) + t(q(1, X) - q_0(1, X)))^3 \right) \right],
\end{aligned}$$

$$\begin{aligned}
\mathcal{J}_2 &:= -2\mathbb{E} \left[ (1 - D) R \left\{ (g(0, X) - g_0(0, X)) (1 - \mathbb{E}[D] + t(\mathbb{E}[D] - p_D)) \right. \right. \\
&\quad \times (-r_0(0, X) + t(r_0(0, X) - r(0, X))) \\
&\quad \left. \left. \times ((r_0(0, X) - r(0, X)) (1 - \mathbb{E}[D] + t(2\mathbb{E}[D] - 2p_D)) + r_0(0, X) (p_D - \mathbb{E}[D])) \right\} \right]
\end{aligned}$$

$$\begin{aligned}
& + \left( (r(0, X) - r_0(0, X))^2 (1 - E[D] + t(E[D] - p_D))^2 \right. \\
& \quad - (p_D - E[D]) (r(0, X) - r_0(0, X)) (1 - E[D] + t(E[D] - p_D)) \\
& \quad \times (r_0(0, X) + t(r(0, X) - r_0(0, X))) \\
& \quad \left. + (p_D - E[D])^2 (r_0(0, X) + t(r(0, X) - r_0(0, X)))^2 \right) \\
& \quad \times (Y - g_0(0, X) + t(g_0(0, X) - g(0, X))) \Big\} \\
& / \left( (1 - E[D] + t(E[D] - p_D))^3 (r_0(0, X) + t(r(0, X) - r_0(0, X)))^3 \right) \Big], \\
& \mathcal{J}_3 := -2E \left[ (1 - D) (p_D - E[D]) \left( (g_0(0, X) - g_0(1, X)) (p_D - E[D]) \right. \right. \\
& \quad \left. \left. + (g(0, X) - g_0(0, X) + g_0(1, X) - g(1, X)) (1 - E[D]) \right) \right. \\
& \quad \left. / (1 - E[D] + t(E[D] - p_D))^3 \right], \\
& \mathcal{J}_4 := 2E \left[ (1 - D) \gamma_0 (p_D - E[D])^2 / (1 - E[D] + t(E[D] - p_D))^3 \right].
\end{aligned}$$

By assumption we have  $DR(Y - g(1, X)) = DRU$ ,

$$(1 - D)R(Y - g_0(0, X)) = (1 - D)RU, \quad E[U|D, R, X] = 0,$$

$$\|g(D, X) - g_0(D, X)\|_2 \times \|m(X) - m_0(X)\|_2 \leq \delta_N N^{-1/2},$$

$$\|g(D, X) - g_0(D, X)\|_2 \times \|r(D, X) - r_0(D, X)\|_2 \leq \delta_N N^{-1/2},$$

$$\|g(D, X) - g_0(D, X)\|_2 \times \|q(D, X) - q_0(D, X)\|_2 \leq \delta_N N^{-1/2}.$$

In addition,  $\forall i \in \{1, \dots, l\} : \|p_D - E[D]\|_2 \times \|\ell_i - \ell_{0,i}\|_2 \leq \delta_N N^{-1/2}$  since the sample mean estimator is independent of the other nuisance functions and (at least)  $o(N^{-1/2})$ .

Putting these together, and using the Cauchy–Bunyakovsky–Schwarz inequality, it follows that

$$\left| \frac{\partial^2}{\partial t^2} E[\varphi_\gamma(W; \gamma_0, \eta_0 + t(\eta - \eta_0))] \right| \leq C' \delta_N / \sqrt{N},$$

with constant  $C'$  depending only on  $\varepsilon$ ,  $\gamma_0$ , and  $p_{\max}$ . This concludes the proof.  $\square$

### 2.A.5 Data Preparation and Hyperparameter Tuning

For the random forest, boosted trees, and SVM, the dictionary of considered controls encompasses all the variables listed in Table 2.A.1, with categorical variables expanded into dummy variables. For the regularized regression techniques, I use an extended set of can-

didate variables. Next to the variables already mentioned, it includes squared and cubed terms of all numerical variables, and cubic B-splines with five interior knots of three continuous variables—two household poverty indices and a village-level poverty index. Last of all, it includes interactions of all the previously mentioned variables with a subset of 28 variables that are deemed particularly relevant; these include student characteristics, household demographics, parents’ education and expectations, local wages paid to children, and poverty levels. Missing values are treated as follows: for categorical variables, a new *missing* category is created. For numerical variables, missing entries are assigned the average of all non-missing entries, and an additional *missing* dummy is created. Extreme gradient boosting is the only method which does not require missing value imputation. After dropping duplicates and perfectly collinear variables, the basic dictionary of variables for the eligible students in the conditional sample includes 222 variables, whereas the extended dictionary includes 16,395 variables. The sets of variables for the non-eligible students, as well as for eligible and non-eligible students in the unconditional sample, are very similar in magnitude.

To obtain the best possible prediction model for each machine learning method and nuisance function, a number of hyperparameters need to be selected. For the regularized regression techniques, these include the  $\ell_1$  and  $\ell_2$  regularization parameters. For SVM, they are the cost as well as the parameter  $\gamma$  for the radial basis kernel. For boosting with logistic regression trees, the parameters are number of boosting iterations, learning rate, maximal tree depth, minimum loss reduction, subsample ratio of training set observations, subsample ratio of variables, and minimum sum of instance weight per leaf. For random forests, where overfitting is not a concern, I go without hyperparameter tuning and simply choose a large enough number of trees (1,000) and a leaf size of 1.

For each nuisance function and method, I create a grid with likely values for the hyperparameters and run a repeated cross-validation. To that end, the dataset is split in the same way as for the cross-fitting procedure, i.e., in 10 folds, with roughly equal ratios of treated observations, and with no overlapping locations. For each hyperparameter vector from the grid, the model is tuned in 9 folds and predictions made in the remaining fold. This is done separately for all 10 folds, and repeated a total of 10 times for different random splits. In the end, the hyperparameter vector with the lowest average out-of-sample mean squared error is selected.

### 2.A.6 Result Tables

Table 2.A.2: Detailed estimates of conditional ATE and ATN on high school education for adolescents from eligible (poor) households; unrestricted sample

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Best method
<i>(1) Dependent variable: going to high school in November 2000.</i>							
ATE	-0.124** (0.055)	-0.107* (0.064)	-0.127** (0.054)	-0.123** (0.062)	-0.107 (0.065)	-0.127* (0.077)	-0.125** (0.055)
ATN	-0.111** (0.056)	-0.106 (0.065)	-0.110** (0.055)	-0.116* (0.061)	-0.113* (0.064)	-0.109 (0.084)	-0.107* (0.057)
<i>(2) Dependent variable: some high school by 2003.</i>							
ATE	-0.101* (0.060)	-0.076 (0.065)	-0.101* (0.059)	-0.100 (0.065)	-0.070 (0.089)	-0.077 (0.065)	-0.101* (0.059)
ATN	-0.061 (0.068)	-0.076 (0.065)	-0.064 (0.067)	-0.068 (0.064)	-0.041 (0.082)	-0.046 (0.069)	-0.067 (0.066)
<i>(3) Dependent variable: graduated or about to graduate from high school in 2003.</i>							
ATE	-0.092 (0.057)	-0.074 (0.063)	-0.099* (0.056)	-0.114* (0.064)	-0.058 (0.082)	-0.075 (0.063)	-0.100* (0.056)
ATN	-0.082 (0.059)	-0.073 (0.064)	-0.086 (0.059)	-0.091 (0.064)	-0.051 (0.079)	-0.065 (0.068)	-0.091 (0.058)

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

Average treatment effects (ATE) and average treatment effects of the non-treated (ATN) as well as standard errors (in parentheses) are estimated for adolescents from eligible households in the unrestricted conditional sample. That includes adolescents whose highest completed grade by November 2000 was the third grade of middle school, and who had been enrolled in school during the entire academic year 1999/2000. Column labels denote the method used to estimate the nuisance functions.

Point estimates for the first outcome are obtained via the orthogonal moment conditions of equations (2.A.1) and (2.A.2), while their respective modifications for attrition, (2.A.3) and (2.A.4), are employed for the other two outcomes, as these are unobserved for some adolescents. Standard errors are obtained via wild cluster bootstrap. Results are based on 10-fold cross-fitting, using 100 sample splits. Estimates from the sample splits are combined using the median method, based on equations (2.A.6) and (2.A.7).



Table 2.A.3: Detailed estimates of conditional ATE and ATN on high school education for adolescents from eligible (poor) households; restricted sample

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Best method
<i>(1) Dependent variable: going to high school in November 2000.</i>							
ATE	-0.141** (0.069)	-0.135* (0.073)	-0.146** (0.069)	-0.166** (0.076)	-0.171** (0.074)	-0.147** (0.070)	-0.145** (0.069)
ATN	-0.124* (0.070)	-0.135* (0.073)	-0.139** (0.069)	-0.155** (0.079)	-0.165** (0.082)	-0.140** (0.070)	-0.135* (0.070)
<i>(2) Dependent variable: some high school by 2003.</i>							
ATE	-0.119 (0.078)	-0.113 (0.078)	-0.122 (0.078)	-0.172** (0.084)	-0.106 (0.158)	-0.144* (0.082)	-0.126 (0.078)
ATN	-0.120 (0.078)	-0.114 (0.079)	-0.127 (0.078)	-0.125 (0.087)	-0.219 (0.144)	-0.145* (0.081)	-0.126 (0.077)
<i>(3) Dependent variable: graduated or about to graduate from high school in 2003.</i>							
ATE	-0.092 (0.076)	-0.086 (0.077)	-0.091 (0.077)	-0.146* (0.085)	-0.046 (0.174)	-0.119 (0.081)	-0.098 (0.076)
ATN	-0.092 (0.076)	-0.086 (0.077)	-0.092 (0.077)	-0.096 (0.086)	-0.162 (0.153)	-0.114 (0.078)	-0.099 (0.076)

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

Average treatment effects (ATE) and average treatment effects of the non-treated (ATN) as well as standard errors (in parentheses) are estimated for adolescents from eligible households in the restricted conditional sample. That includes adolescents whose highest completed grade by November 1999 was the second grade of middle school, by November 2000 the third grade of middle school, and who had been enrolled in school during the entire academic year 1999/2000. Column labels denote the method used to estimate the nuisance functions.

Point estimates for the first outcome are obtained via the orthogonal moment conditions of equations (2.A.1) and (2.A.2), while their respective modifications for attrition, (2.A.3) and (2.A.4), are employed for the other two outcomes, as these are unobserved for some adolescents. Standard errors are obtained via wild cluster bootstrap. Results are based on 10-fold cross-fitting, using 100 sample splits. Estimates from the sample splits are combined using the median method, based on equations (2.A.6) and (2.A.7).

Table 2.A.4: Detailed estimates of conditional ATE and ATN on high school education for adolescents from non-eligible (non-poor) households; unrestricted sample

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Best method
<i>(1) Dependent variable: going to high school in November 2000.</i>							
ATE	0.064 (0.056)	0.023 (0.062)	0.059 (0.055)	0.090 (0.064)	0.104* (0.063)	0.116 (0.075)	0.066 (0.055)
ATN	0.078 (0.054)	0.060 (0.061)	0.077 (0.054)	0.098* (0.056)	0.113* (0.062)	0.115 (0.078)	0.089 (0.056)
<i>(2) Dependent variable: some high school by 2003.</i>							
ATE	0.122* (0.071)	0.086 (0.075)	0.120* (0.071)	0.087 (0.079)	0.103 (0.089)	0.089 (0.080)	0.109 (0.072)
ATN	0.124* (0.070)	0.086 (0.074)	0.121* (0.070)	0.117 (0.075)	0.135 (0.091)	0.093 (0.074)	0.118 (0.074)
<i>(3) Dependent variable: graduated or about to graduate from high school in 2003.</i>							
ATE	0.169*** (0.064)	0.105 (0.067)	0.165*** (0.064)	0.128* (0.070)	0.162* (0.086)	0.142** (0.072)	0.149** (0.064)
ATN	0.167** (0.066)	0.104 (0.069)	0.161** (0.065)	0.132*** (0.067)	0.162 (0.109)	0.123* (0.070)	0.141** (0.071)

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

Average treatment effects (ATE) and average treatment effects of the non-treated (ATN) as well as standard errors (in parentheses) are estimated for adolescents from non-eligible households in the unrestricted conditional sample. That includes adolescents whose highest completed grade by November 2000 was the third grade of middle school, and who had been enrolled in school during the entire academic year 1999/2000. Column labels denote the method used to estimate the nuisance functions.

Point estimates for the first outcome are obtained via the orthogonal moment conditions of equations (2.A.1) and (2.A.2), while their respective modifications for attrition, (2.A.3) and (2.A.4), are employed for the other two outcomes, as these are unobserved for some adolescents. Standard errors are obtained via wild cluster bootstrap. Results are based on 10-fold cross-fitting, using 100 sample splits. Estimates from the sample splits are combined using the median method, based on equations (2.A.6) and (2.A.7).

Table 2.A.5: Detailed estimates of conditional ATE and ATN on high school education for adolescents from non-eligible (non-poor) households; restricted sample

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Best method
<i>(1) Dependent variable: going to high school in November 2000.</i>							
ATE	0.106 (0.073)	0.108 (0.076)	0.102 (0.072)	0.093 (0.074)	0.007 (0.117)	0.090 (0.078)	0.093 (0.073)
ATN	0.103 (0.074)	0.108 (0.077)	0.110 (0.074)	0.112 (0.075)	0.001 (0.139)	0.131 (0.081)	0.110 (0.075)
<i>(2) Dependent variable: some high school by 2003.</i>							
ATE	0.181** (0.084)	0.161* (0.085)	0.169** (0.083)	0.152* (0.083)	0.133 (0.195)	0.157* (0.087)	0.162** (0.080)
ATN	0.182** (0.089)	0.161* (0.086)	0.177** (0.088)	0.186** (0.089)	0.204 (0.160)	0.151* (0.090)	0.184** (0.084)
<i>(3) Dependent variable: graduated or about to graduate from high school in 2003.</i>							
ATE	0.216*** (0.079)	0.194** (0.080)	0.209*** (0.079)	0.182** (0.078)	0.180 (0.203)	0.193** (0.080)	0.193** (0.078)
ATN	0.212** (0.082)	0.195** (0.081)	0.213** (0.084)	0.200** (0.084)	0.220 (0.169)	0.177** (0.088)	0.191** (0.086)

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

Average treatment effects (ATE) and average treatment effects of the non-treated (ATN) as well as standard errors (in parentheses) are estimated for adolescents from non-eligible households in the restricted conditional sample. That includes adolescents whose highest completed grade by November 1999 was the second grade of middle school, by November 2000 the third grade of middle school, and who had been enrolled in school during the entire academic year 1999/2000. Column labels denote the method used to estimate the nuisance functions.

Point estimates for the first outcome are obtained via the orthogonal moment conditions of equations (2.A.1) and (2.A.2), while their respective modifications for attrition, (2.A.3) and (2.A.4), are employed for the other two outcomes, as these are unobserved for some adolescents. Standard errors are obtained via wild cluster bootstrap. Results are based on 10-fold cross-fitting, using 100 sample splits. Estimates from the sample splits are combined using the median method, based on equations (2.A.6) and (2.A.7).

Table 2.A.6: Detailed estimates of unconditional ATE on high school education for adolescents from eligible (poor) households

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Best method
<i>(1) Dependent variable: going to high school in November 2000.</i>							
ATE	-0.033	-0.030	-0.033	-0.042	-0.031	-0.009	-0.033
	(0.030)	(0.031)	(0.030)	(0.031)	(0.050)	(0.033)	(0.030)
<i>(2) Dependent variable: some high school by 2003.</i>							
ATE	-0.091**	-0.075	-0.089**	-0.080*	-0.067	-0.074	-0.085**
	(0.044)	(0.047)	(0.044)	(0.043)	(0.041)	(0.056)	(0.043)
<i>(3) Dependent variable: graduated or about to graduate from high school in 2003.</i>							
ATE	-0.046	-0.050	-0.045	-0.062	-0.046	-0.041	-0.050
	(0.036)	(0.040)	(0.036)	(0.039)	(0.034)	(0.047)	(0.035)

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
Average treatment effects (ATE) and average treatment effects of the non-treated (ATN) as well as standard errors (in parentheses) are estimated for adolescents from eligible households in the unconditional sample. That includes adolescents who had graduated from primary school in 1997. Column labels denote the method used to estimate the nuisance functions.  
Point estimates are obtained via the orthogonal moment conditions of the equations that account for attrition, (2.A.3) and (2.A.4). Standard errors are obtained via wild cluster bootstrap. Results are based on 10-fold cross-fitting, using 100 sample splits. Estimates from the sample splits are combined using the median method, based on equations (2.A.6) and (2.A.7).

Table 2.A.7: Detailed estimates of unconditional ATE on high school education for adolescents from non-eligible (non-poor) households

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Best method
<i>(1) Dependent variable: going to high school in November 2000.</i>							
ATE	0.124*** (0.034)	0.123*** (0.036)	0.124*** (0.034)	0.127*** (0.034)	0.145* (0.077)	0.139*** (0.038)	0.129*** (0.035)
<i>(2) Dependent variable: some high school by 2003.</i>							
ATE	0.093* (0.048)	0.115** (0.050)	0.089* (0.047)	0.094** (0.046)	0.087 (0.068)	0.091 (0.061)	0.100** (0.044)
<i>(3) Dependent variable: graduated or about to graduate from high school in 2003.</i>							
ATE	0.101** (0.042)	0.135*** (0.042)	0.098** (0.041)	0.105*** (0.040)	0.108* (0.059)	0.101* (0.056)	0.107*** (0.038)

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
Average treatment effects (ATE) and average treatment effects of the non-treated (ATN) as well as standard errors (in parentheses) are estimated for adolescents from non-eligible households in the unconditional sample. That includes adolescents who had graduated from primary school in 1997. Column labels denote the method used to estimate the nuisance functions.  
Point estimates are obtained via the orthogonal moment conditions of the equations that account for attrition, (2.A.3) and (2.A.4). Standard errors are obtained via wild cluster bootstrap. Results are based on 10-fold cross-fitting, using 100 sample splits. Estimates from the sample splits are combined using the median method, based on equations (2.A.6) and (2.A.7).

Table 2.A.8: Detailed estimates of unconditional ATE on middle school education for adolescents from eligible (poor) households

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Best method
<i>(1) Dependent variable: graduated from middle school by 2000.</i>							
ATE	-0.014	-0.025	-0.013	-0.039	-0.072	-0.043	-0.029
	(0.032)	(0.037)	(0.032)	(0.036)	(0.064)	(0.049)	(0.032)
<i>(2) Dependent variable: graduated or about to graduate from middle school in 2000.</i>							
ATE	0.089***	0.087**	0.090***	0.068*	0.031	0.052	0.088***
	(0.031)	(0.037)	(0.031)	(0.035)	(0.069)	(0.045)	(0.032)

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
Average treatment effects (ATE) and average treatment effects of the non-treated (ATN) as well as standard errors (in parentheses) are estimated for adolescents from eligible households in the unconditional sample. That includes adolescents who had graduated from primary school in 1997. Column labels denote the method used to estimate the nuisance functions.  
Point estimates are obtained via the orthogonal moment conditions of the equations that account for attrition, (2.A.3) and (2.A.4). Standard errors are obtained via wild cluster bootstrap. Results are based on 10-fold cross-fitting, using 100 sample splits. Estimates from the sample splits are combined using the median method, based on equations (2.A.6) and (2.A.7).

Table 2.A.9: Detailed estimates of unconditional ATE on middle school education for adolescents from non-eligible (non-poor) households

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Best method
<i>(1) Dependent variable: graduated from middle school by 2000.</i>							
ATE	0.063 (0.044)	0.066 (0.050)	0.068 (0.044)	0.079* (0.043)	0.114 (0.088)	0.094 (0.051)	0.069 (0.044)
<i>(2) Dependent variable: graduated or about to graduate from middle school in 2000.</i>							
ATE	0.023 (0.036)	0.042 (0.045)	0.024 (0.036)	0.055 (0.039)	0.092 (0.081)	0.086* (0.049)	0.026 (0.037)

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
Average treatment effects (ATE) and average treatment effects of the non-treated (ATN) as well as standard errors (in parentheses) are estimated for adolescents from non-eligible households in the unconditional sample. That includes adolescents who had graduated from primary school in 1997. Column labels denote the method used to estimate the nuisance functions.  
Point estimates are obtained via the orthogonal moment conditions of the equations that account for attrition, (2.A.3) and (2.A.4). Standard errors are obtained via wild cluster bootstrap. Results are based on 10-fold cross-fitting, using 100 sample splits. Estimates from the sample splits are combined using the median method, based on equations (2.A.6) and (2.A.7).

Table 2.A.10: Detailed estimates of conditional ATE and ATN on further outcomes for adolescents from eligible (poor) households; unrestricted sample

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Best method
<i>(1) Dependent variable: not going to high school due to financial constraints.</i>							
ATE	0.045 (0.054)	0.032 (0.057)	0.048 (0.053)	0.031 (0.058)	0.031 (0.061)	0.050 (0.081)	0.035 (0.053)
ATN	0.036 (0.057)	0.031 (0.057)	0.036 (0.056)	0.002 (0.060)	0.017 (0.073)	0.015 (0.093)	0.019 (0.055)
<i>(2) Dependent variable: not going to high school for other than financial reasons.</i>							
ATE	0.071* (0.043)	0.075* (0.042)	0.073* (0.042)	0.079* (0.045)	0.075 (0.047)	0.075 (0.064)	0.073* (0.042)
ATN	0.074* (0.044)	0.075* (0.044)	0.074* (0.044)	0.089* (0.051)	0.077 (0.052)	0.098 (0.083)	0.074* (0.044)
<i>(3) Dependent variable: log monthly per capita expenditure.</i>							
ATE	0.051 (0.066)	0.059 (0.068)	0.051 (0.066)	0.051 (0.071)	0.043 (0.070)	0.060 (0.066)	0.059 (0.063)
ATN	0.052 (0.063)	0.058 (0.068)	0.052 (0.063)	0.043 (0.068)	0.045 (0.073)	0.060 (0.066)	0.051 (0.063)

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
Average treatment effects (ATE) and average treatment effects of the non-treated (ATN) as well as standard errors (in parentheses) are estimated for adolescents from eligible households in the unrestricted conditional sample. That includes adolescents whose highest completed grade by November 2000 was the third grade of middle school, and who had been enrolled in school during the entire academic year 1999/2000. Column labels denote the method used to estimate the nuisance functions.  
Point estimates are obtained via the orthogonal moment conditions of equations (2.A.1) and (2.A.2). Standard errors are obtained via wild cluster bootstrap. Results are based on 10-fold cross-fitting, using 100 sample splits. Estimates from the sample splits are combined using the median method, based on equations (2.A.6) and (2.A.7).



Table 2.A.11: Detailed estimates of conditional ATE and ATN on further outcomes for adolescents from eligible (poor) households; restricted sample

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Best method
<i>(1) Dependent variable: not going to high school due to financial constraints.</i>							
ATE	0.031 (0.064)	0.038 (0.065)	0.031 (0.064)	0.040 (0.068)	0.069 (0.072)	0.038 (0.065)	0.031 (0.064)
ATN	0.044 (0.064)	0.039 (0.065)	0.044 (0.064)	0.035 (0.074)	0.059 (0.093)	0.038 (0.065)	0.044 (0.064)
<i>(2) Dependent variable: not going to high school for other than financial reasons.</i>							
ATE	0.109** (0.052)	0.097* (0.054)	0.114** (0.051)	0.106* (0.055)	0.082 (0.059)	0.104* (0.054)	0.113** (0.051)
ATN	0.103* (0.056)	0.096* (0.055)	0.103* (0.056)	0.090 (0.063)	0.040 (0.085)	0.087 (0.058)	0.103* (0.056)
<i>(3) Dependent variable: log monthly per capita expenditure.</i>							
ATE	0.019 (0.078)	0.017 (0.080)	0.016 (0.079)	0.020 (0.086)	0.031 (0.082)	0.018 (0.080)	0.019 (0.077)
ATN	0.023 (0.076)	0.017 (0.079)	0.023 (0.076)	-0.016 (0.082)	0.026 (0.082)	0.018 (0.079)	0.016 (0.075)

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

Average treatment effects (ATE) and average treatment effects of the non-treated (ATN) as well as standard errors (in parentheses) are estimated for adolescents from eligible households in the restricted conditional sample. That includes adolescents whose highest completed grade by November 1999 was the second grade of middle school, by November 2000 the third grade of middle school, and who had been enrolled in school during the entire academic year 1999/2000. Column labels denote the method used to estimate the nuisance functions.

Point estimates are obtained via the orthogonal moment conditions of equations (2.A.1) and (2.A.2). Standard errors are obtained via wild cluster bootstrap. Results are based on 10-fold cross-fitting, using 100 sample splits. Estimates from the sample splits are combined using the median method, based on equations (2.A.6) and (2.A.7).

Table 2.A.12: Detailed estimates of conditional ATE and ATN on further outcomes for adolescents from non-eligible (non-poor) households; unrestricted sample

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Best method
<i>(1) Dependent variable: not going to high school due to financial constraints.</i>							
ATE	0.000 (0.059)	0.017 (0.061)	0.002 (0.058)	-0.013 (0.061)	-0.020 (0.063)	-0.001 (0.077)	0.002 (0.057)
ATN	-0.021 (0.057)	0.017 (0.061)	-0.015 (0.056)	-0.031 (0.053)	-0.046 (0.059)	0.002 (0.079)	-0.014 (0.055)
<i>(2) Dependent variable: not going to high school for other than financial reasons.</i>							
ATE	-0.087* (0.046)	-0.078* (0.047)	-0.087* (0.046)	-0.076 (0.050)	-0.064 (0.052)	-0.106* (0.061)	-0.087* (0.046)
ATN	-0.084* (0.047)	-0.076 (0.048)	-0.085* (0.047)	-0.075 (0.048)	-0.080 (0.052)	-0.120** (0.054)	-0.085* (0.047)
<i>(3) Dependent variable: log monthly per capita expenditure.</i>							
ATE	-0.090 (0.086)	-0.107 (0.088)	-0.090 (0.085)	-0.067 (0.091)	-0.076 (0.092)	-0.105 (0.087)	-0.079 (0.081)
ATN	-0.113 (0.086)	-0.107 (0.089)	-0.112 (0.086)	-0.084 (0.085)	-0.082 (0.094)	-0.107 (0.088)	-0.104 (0.085)

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
Average treatment effects (ATE) and average treatment effects of the non-treated (ATN) as well as standard errors (in parentheses) are estimated for adolescents from non-eligible households in the unrestricted conditional sample. That includes adolescents whose highest completed grade by November 2000 was the third grade of middle school, and who had been enrolled in school during the entire academic year 1999/2000. Column labels denote the method used to estimate the nuisance functions.  
Point estimates are obtained via the orthogonal moment conditions of equations (2.A.1) and (2.A.2). Standard errors are obtained via wild cluster bootstrap. Results are based on 10-fold cross-fitting, using 100 sample splits. Estimates from the sample splits are combined using the median method, based on equations (2.A.6) and (2.A.7).

Table 2.A.13: Detailed estimates of conditional ATE and ATN on further outcomes for adolescents from non-eligible (non-poor) households; restricted sample

	Lasso	Ridge	Elastic net	Random forest	Boosted trees	SVM	Best method
<i>(1) Dependent variable: not going to high school due to financial constraints.</i>							
ATE	-0.018 (0.070)	-0.020 (0.073)	-0.013 (0.070)	-0.009 (0.074)	0.090 (0.118)	-0.012 (0.076)	-0.010 (0.072)
ATN	-0.015 (0.068)	-0.021 (0.073)	-0.019 (0.070)	-0.038 (0.069)	0.052 (0.144)	-0.039 (0.077)	-0.037 (0.069)
<i>(2) Dependent variable: not going to high school for other than financial reasons.</i>							
ATE	-0.088 (0.061)	-0.088 (0.061)	-0.088 (0.060)	-0.091 (0.062)	-0.095 (0.098)	-0.086 (0.063)	-0.085 (0.060)
ATN	-0.087 (0.061)	-0.087 (0.061)	-0.088 (0.061)	-0.083 (0.060)	-0.080 (0.108)	-0.091 (0.064)	-0.085 (0.062)
<i>(3) Dependent variable: log monthly per capita expenditure.</i>							
ATE	-0.133 (0.098)	-0.160 (0.101)	-0.138 (0.098)	-0.103 (0.102)	-0.132 (0.166)	-0.162 (0.101)	-0.113 (0.101)
ATN	-0.145 (0.097)	-0.158 (0.104)	-0.148 (0.097)	-0.139 (0.098)	-0.148 (0.205)	-0.159 (0.105)	-0.139 (0.096)

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

Average treatment effects (ATE) and average treatment effects of the non-treated (ATN) as well as standard errors (in parentheses) are estimated for adolescents from non-eligible households in the restricted conditional sample. That includes adolescents whose highest completed grade by November 1999 was the second grade of middle school, by November 2000 the third grade of middle school, and who had been enrolled in school during the entire academic year 1999/2000. Column labels denote the method used to estimate the nuisance functions.

Point estimates are obtained via the orthogonal moment conditions of equations (2.A.1) and (2.A.2). Standard errors are obtained via wild cluster bootstrap. Results are based on 10-fold cross-fitting, using 100 sample splits. Estimates from the sample splits are combined using the median method, based on equations (2.A.6) and (2.A.7).

### 2.A.7 Omitted Variable Bias Estimation

In this section, I compute an estimator for the omitted variable bias by Oster (2017). Its basis is an outcome model of the form

$$Y = D\theta + Z_1 + Z_2 + \epsilon, \quad (2.A.11)$$

with an observed part  $Z_1 = X\beta$  and an unobserved part  $Z_2$ . This model is more restrictive than the heterogeneous treatment effect model, most notably since both treatment status  $D$  and observables  $X$  enter the outcome model linearly. This means that the results presented here should merely be seen as back-of-the-envelope calculations. Let  $\hat{\theta}$  be the point estimate of the estimation excluding the unobservable part, and let  $\hat{\theta}^\circ$  be the point estimate of a simple regression of  $Y$  on  $D$ . Let  $\hat{R}^\circ$ ,  $\hat{R}$ , and  $R_{\max}$  respectively denote the  $R$ -squared of the simple regression model, of the model with observed characteristics, and of the (hypothetical) outcome model with all observed and unobserved pre-treatment variables included. Furthermore, let  $\delta$  denote the relative importance of selection on the observed and the unobserved part of the model, i.e.,

$$\delta \frac{\text{cov}(Z_1, D)}{\text{var}(Z_1)} = \frac{\text{cov}(Z_2, D)}{\text{var}(Z_2)}.$$

The omitted variable bias  $\Pi$  is then estimated as

$$\hat{\Pi} = \delta \left( \hat{\theta}^\circ - \hat{\theta} \right) \left( R_{\max} - \hat{R} \right) / \left( \hat{R} - \hat{R}^\circ \right). \quad (2.A.12)$$

Among the components in equation (2.A.12),  $\hat{\theta}$ ,  $\hat{\theta}^\circ$ ,  $\hat{R}$ , and  $\hat{R}^\circ$  are observed, while  $R_{\max}$  and  $\delta$  are not.  $R_{\max}$  represents the maximal  $R$ -squared achievable in a prediction of  $Y$  using pre-treatment information, a value bounded from above by 1. It is reasonable to expect  $R_{\max}$  to be lower than 1, for the following two reasons. First, all outcomes are measured years after the time before treatment. Therefore, idiosyncratic shocks can occur after the beginning of the treatment, which affect outcomes but are by definition unpredictable using pre-treatment information. Second, any measurement error in  $Y$  reduces predictability. For this exercise, I chose multiples of  $\hat{R}$  as possible values for  $R_{\max}$ , with multipliers 2, 3, 4, and 5. I then calculate the value of  $\delta$  for which the true treatment effect would be zero, or  $\hat{\Pi} = \hat{\theta}$ . A value for  $\delta$  that is far from 0 is unlikely—given the effort to include all covariates that are highly correlated with  $D$ —and in turn means a null result is unlikely. Oster argues that  $\delta = 1$  is an appropriate cutoff value, as this implies that unobservables are as important as observables. I adopt this argument, with

the qualification that  $\text{cov}(Z_1, D)$  and  $\text{cov}(Z_2, D)$  may have opposite signs. This means that for any given  $R_{\max}$ ,  $\theta = 0$  is rejected if  $|\delta| > 1$ .

Table 2.A.14 shows the  $\delta$  implied by  $\theta = 0$ , where  $\theta$  denotes the true average treatment effect, for  $R_{\max} = \min(i\hat{R}, 1)$  with  $i = 2, \dots, 5$  and all samples and school-related outcomes discussed in the chapter. It can be seen that for all results that are at least significant at the 10% level,  $|\delta| > 1$  even if the unobserved variables explain three times as much as the observed ones, i.e., if  $R_{\max} = 4\hat{R}$ . A small caveat here is that the values of  $\hat{\theta}$  do not take account of observations with missing outcomes, whereas the values of  $\hat{\theta}$  do. Thus, for samples with some outcomes missing, the estimates of  $\delta$  are likely still too close to 0. For all those samples without missing observations, it holds that  $|\delta| > 1$  even if  $R_{\max} = 5\hat{R}$ . In conclusion, even a significant amount of unobserved pre-treatment information is likely not going to invalidate the main results.

Table 2.A.14: Relative importance of selection on the observed and the unobserved

$R_{\max} = \min(\cdot, 1)$	eligible (poor)					non-eligible (non-poor)				
	$2\hat{R}$	$3\hat{R}$	$4\hat{R}$	$5\hat{R}$		$2\hat{R}$	$3\hat{R}$	$4\hat{R}$	$5\hat{R}$	
<i>Unconditional sample</i>										
Going to high school in November 2000	-12.42	-6.21	-4.14	-3.11		-58.0	-29.0	-19.3	-14.5	
Some high school by 2003	-8.57	-4.28	-2.86	-2.14		-8.48	-4.24	-2.83	-2.12	
Graduated or about to graduate from high school in 2003	125.3	62.7	41.8	31.3		9.77	4.88	3.26	2.44	
Graduated from middle school by 2000	-2.28	-1.14	-0.76	-0.71		-143.0	-71.5	-47.7	-35.7	
Graduated or about to graduate from middle school in 2000	-63.9	-31.94	-21.3	-18.4		1.58	0.79	0.53	0.47	
<i>Conditional unrestricted sample</i>										
Going to high school in November 2000	-6.29	-3.15	-2.10	-1.57		-10.2	-5.11	-3.41	-2.56	
Some high school by 2003	-3.80	-1.90	-1.27	-0.95		-3.94	-1.97	-1.31	-0.98	
Graduated or about to graduate from high school in 2003	-3.45	-1.73	-1.15	-0.86		-3.18	-1.59	-1.06	-0.79	
<i>Conditional restricted sample</i>										
Going to high school in November 2000	-10.6	-5.29	-3.53	-2.64		5.91	2.96	1.97	1.48	
Some high school by 2003	-2.22	-1.11	-0.74	-0.56		-51.4	-25.7	-17.1	-12.9	
Graduated or about to graduate from high school in 2003	-2.47	-1.24	-0.82	-0.62		249.6	124.8	83.2	62.4	

# Chapter 3

## Welfare Measurement and Poverty Targeting Based on Participatory Wealth Rankings

Social security programs that target the poor are staple policies in many developing countries. Where tax bases are shallow and incidences of poverty high, selecting beneficiaries is often preferred to universal coverage. The problem of how best to choose who should be eligible is subject of ongoing discourse. A substantial branch of literature is concerned with the question how well different targeting methods, e.g. proxy means testing, geographical targeting, participatory methods, or self-selection, align with certain targeting objectives, usually poverty status as measured by consumption or income (Coady, Grosh and Hoddinott, 2004; Zeller, Feulefack and Neef, 2006; Banerjee et al., 2009; Coady and Parker, 2009; Yusuf, 2010; Alatas et al., 2012; Alatas et al., 2016; Alatas et al., 2019; Bah et al., 2019; Karlan and Thuysbaert, 2019). A related yet less explored question is how the method of targeting affects acceptance and satisfaction with the programs in question.<sup>1</sup> How well a targeted intervention is being received may depend on a number of components: the degree to which the allocation meets shared intuitions of justice, the amount of self-determination allowed to communities in the targeting process, as well as

---

<sup>1</sup>The significance of dissatisfaction due to targeting and program implementation can hardly be overstated. The Indonesian Direct Cash Assistance (Bantuan Langsung Tunai, or BLT) programs launched in 2005 and 2008 illustrate this point. For the implementation of 2005, Widjaja (2012) reports nationwide protests, threats to staff of the Central Statistics Bureau, and cases of vandalism against government facilities as a result of deficient program implementation. Cameron and Shah (2014) find for the same program that an increase in crime and a decrease in people's participation in community groups are associated with mistargeting. Alatas et al. (2012) remark that for the 2008 implementation of the BLT program, dissatisfaction with beneficiary lists was so immense that more than 2000 village officials refused to participate in the program.

the extent to which the specific needs of different localities are met.

This chapter is an attempt to formalize these factors, and to identify how they impact program satisfaction in the context of a field experiment on targeting in Indonesia, conducted by Alatas et al. (2012). To understand what drives satisfaction with antipoverty programs, it is essential to take into account local views of what constitutes poverty. One way of doing so is through participatory approaches, where data collection and targeting come with an active involvement of the local population. The prime example of such approaches is the participatory wealth ranking (PWR): representatives of a community rank all households according to their wealth. Such rankings can help to understand how poverty is perceived by locals, and how different household characteristics are weighted in the assessment. An important insight from PWRs is that local perceptions of welfare differ systematically from the traditional assessments based on income or consumption (Shaffer, 2013). Furthermore, evidence from field experiments suggests that participants are generally satisfied with the results of interventions that use PWRs as the targeting method (Alatas et al., 2012; Schüring, 2014). Taken together, these findings suggest that PWRs are not only pragmatic ways to allocate benefits at the local level, but that they may also be used as the basis for alternative welfare measures.

The first goal of the chapter is to demonstrate how welfare measures grounded in local concepts of material well-being can be constructed from PWRs. The idea is to estimate the relationship between rankings and household characteristics, and to predict scores based on this model. The resulting welfare measure has a number of desirable properties: it does not depend on preselected dimensions of wealth or deprivation, or on predefined weights<sup>2</sup>; it does not rely on subjective categories of welfare; it can be constructed for localities where no actual PWRs have been conducted; and it can be used to relate households from different communities to each other, and thus overcome the principal incomparability of ranking outcomes between villages. The new welfare scores can in turn be used as targeting goals, or as benchmarks to assess targeting performance and to measure local poverty.

A number of arguments can be made why these scores may be more appropriate welfare indicators than consumption. Figures of consumption are usually constructed on the basis of assets and expenses within the days before data collection. This means that prospects for future consumption, income volatility, and the ability to smooth out shocks, are not fully reflected in this measure—though they might be visible to other locals and get incorporated in the PWRs. Furthermore, an antipoverty program should

---

<sup>2</sup>This distinguishes it from other concepts of multi-dimensional poverty, notably the one by Alkire and Foster (2011).



arguably not distribute benefits to households with the lowest consumption, but rather to those with the highest need for them. Or the aim might be to facilitate yet another welfare goal, such as equality of basic capabilities (Sen, 1980). If villagers share these intuitions for distributive justice, the PWRs may lead to more favorable outcomes than rankings based on consumption. Another attempt to assigning normative validity to the outcomes of community rankings is offered by Kanbur and Shaffer (2007). They view participatory approaches in the light of discourse ethics, according to which norms receive validity through practical discourse—an ideal communicative exchange in which participants engage in rational argumentation, and which allow the fair involvement of everyone.<sup>3</sup>

The second goal of the chapter is to estimate the impact that targeting method, targeting accuracy, and amount of provided benefits have on program satisfaction. I use a dataset from a field experiment in Indonesia (Alatas et al., 2012), in which multiple targeting methods were compared to distribute a one-time lump-sum payout. The experiment contained three treatments regarding the within-village allocation of payments to households: a proxy means test (i.e., a predicted consumption score), a PWR, and a hybrid method between the first two. The authors of the study find that people in the PWR villages were more satisfied than those who were targeted based on a proxy means test. It is not clear, though, whether this difference in satisfaction is a result of PWRs leading to an allocation more in line with people's preferences, or because they grant more agency in the allocation process. After defining a measure of targeting accuracy, the experimental setup allows me to separate the impact of the participatory process and of the degree to which the resulting allocation is aligned with the welfare measure. I find that targeting accuracy based on perceived welfare has clearly positive impacts on various outcomes related to program satisfaction. At the same time, there is hardly any evidence that the ranking exercises themselves impact satisfaction, when controlling for targeting accuracy. The effects on satisfaction of (mis)allocating benefits within and between villages can also be measured separately. The results suggest that, on average, reallocating benefits within a village has a stronger effect on satisfaction than providing additional benefits to the village.

---

<sup>3</sup>In how far this ideal of inclusiveness and equal treatment is being met depends on the design and implementation of the PWR, as well as the cultural context. For the wealth rankings from Indonesia considered in this chapter, Alatas et al. (2012) report that when all households of a community were invited, almost half of them participated. The facilitators who moderated the meetings reported only in 15% of the meetings that a few individuals dominated the discussion about the rankings. This indicates that in the case of this particular field experiment, the resulting rankings were indeed the product of a relatively fair and democratic discourse. It is worthwhile to note, however, that the willingness to participate in communal projects and to contribute to public goods in general may be uniquely high in Indonesia for historical reasons (Mansuri and Rao, 2013, chapter 2).

The satisfaction outcomes—besides being of inherent interest—also represent a yardstick against which to evaluate the adequacy of various welfare measures for targeting. In addition to the measures based on local perceptions, I construct targeting accuracy and local poverty based on per capita consumption and examine how they compare in explaining satisfaction. It turns out that the measures based on perceived welfare have a significantly stronger impact than the ones based on consumption. This holds up even when considering only those villages that never conducted a PWR.

The results confirm that local perceptions of welfare are different from consumption-based welfare, and show that this difference is large enough that choosing one over the other as targeting goal translates into noticeable differences in satisfaction. Furthermore, while understanding local perceptions of welfare is thus important for successful targeting, the participatory process itself seems to matter little, if at all. This is a useful insight especially for contexts in which PWRs may not be feasible.

The chapter continues as follows. Section 3.1 summarizes the field experiment and the corresponding dataset of Alatas et al. (2012). In section 3.2, I outline the baseline model of satisfaction and the construction of the different welfare measures, targeting accuracy, and local poverty rates. The welfare prediction models and impact estimations on satisfaction are discussed in section 3.3, and their results are presented in section 3.4. Section 3.5 concludes.

### 3.1 Summary of the Field Experiment and Data Description

The paper by Alatas et al. (2012), which serves as the starting point and primary data source for this study, describes a field experiment conducted in 640 villages in three provinces of Indonesia: North Sumatra, South Sulawesi, and Central Java. The sample of villages is randomly divided into three treatment groups, in which the beneficiary households of an unconditional one-off cash transfer of 30,000 Indonesian Rupees (around 3 US\$) are determined in different ways.

First, in the *PMT* group, household indicators for consumption were collected by the Central Statistics Bureau (BPS) and composed into a proxy means test (PMT) score, using a key that the government had determined through survey data. Households in each village were then ranked according to the PMT score, and the lowest ranked households would receive the benefit. The number of benefits available for each village was determined using an existing poverty map and the Village Potential Statistics (PODES) dataset of 2008, and based on a PPP2\$ per-day poverty line. Second, in the *Community*

group, representatives of households in the village were invited to participate in a ranking exercise, in which households were ranked from poorest to wealthiest. The poorest households would then receive the benefit. A number of sub-treatments were conducted to elicit whether the composition of participants led to any differences. Third, in the *Hybrid* group, the same ranking was conducted as in the Community group, then the BPS collected data to calculate PMT scores for the lowest ranked households, with the cutoff being 1.5 times the number of available benefits for the village. The households with the lowest PMT scores among them received the benefit.

The authors conducted a survey in all the participating villages and constructed a detailed figure of per capita consumption for a sample of nine households per village. They find that the rank-correlation between consumption and the rankings produced by the treatment is highest for the PMT treatment and lowest for the Community treatment—unsurprisingly, given the PMT score was meant to predict consumption. At the same time, satisfaction with the program is higher for the Community treatment than for the two other treatments. Furthermore, the authors do not find any evidence that different subgroups (local elites, women) ranked households differently, that ethnic or religious minorities were discriminated against, or that local community leaders or their relatives were favored. The authors attribute the differences in rank correlation between the treatments to a local understanding of poverty that differs from the consumption metric. Villagers seem to weight especially those factors that are not decisive for consumption but consumption capacity and the ability to smooth shocks. Households whose head was less educated, widowed, disabled, seriously ill, or spent lots of money on tobacco or alcohol were rated relatively lower conditional on consumption, while those with connections to local elites were ranked relatively higher.

For this chapter, I use a variety of components from the dataset of Alatas et al. (2012).<sup>4</sup> The first component is a baseline survey, which contains detailed household information of nine households—the village head and eight households selected at random—from each village, including the results of the ranking exercise and per capita consumption. The baseline survey contains a total of 5,755 observed households. The second component is the data collected by the BPS to obtain PMT scores. It was obtained for every household in the PMT group and for about 47% of households in the Hybrid group. It contains basic information about household demographics, education, occupation, and housing characteristics. On top of that, information about some easily observable assets was collected, including household appliances, electronic devices, livestock, vehicles, productive machinery, and agricultural land. The BPS data includes 10,718 households for the PMT group

---

<sup>4</sup>Available online under <http://dx.doi.org/10.1257/aer.102.4.1206>.

and 5,129 households for the Hybrid group. Table 3.A.1 in the appendix gives a complete list of all the characteristics and assets that were available in both datasets and could be successfully matched and meaningfully used. The baseline survey and the BPS data are used to estimate the relationship between household characteristics and welfare. In addition, the BPS data is used to construct measures of targeting accuracy and poverty headcount for each village.

The third data component is an endline survey, conducted only in Central Java province, with five out of the eight randomly selected households in each village. The fourth component is another endline survey, conducted with the village leaders from all the sampled villages. The two endline surveys are used to evaluate the impact of the treatments and of targeting performance on satisfaction. They include several questions revolving around satisfaction with the program, which are used to construct the set of outcome variables. Households are being asked whether they are satisfied with the program (on a scale of 1 to 4), and village heads are asked whether they think the people in the village are satisfied with the program. Furthermore, village heads and households are asked whether there are any poor households not covered by the program, whether there are any households on the list of beneficiaries who do not belong there, whether the targeting method is correct, whether targeting is worse, equal, or better than the method formerly used for the Direct Cash Assistance (Bantuan Langsung Tunai, or BLT) programs, and if there were too few, enough, or too many benefits given out in the village. For all these questions, I treat answers such as “don’t know” or “no opinion” as missing. In addition, village heads were asked how many complaints about the list of beneficiaries they received. Lastly, there was also a letter box for anonymous complaints, from which the number of complaints is documented.<sup>5</sup> The household and village level outcome variables are summarized in Tables 3.1.1 and 3.1.2, respectively. These tables display a clear pattern: the Community treatment group has significantly better<sup>6</sup> outcomes than the Hybrid and the PMT group, with only a few of the outcomes being statistically indistinguishable. Similarly, the Hybrid group has either significantly better outcomes than the PMT group (mostly for the outcomes at the village level) or is statistically indistinguishable (mostly for the outcomes at the household level).

---

<sup>5</sup>The numbers of complaints are divided by the number of households in the village to make them comparable. The resulting complaints per household are highly skewed, with very few villages registering a lot of complaints. In order to avoid results being driven by a few extreme values, I use a log-transformation of the complaint variables in the regressions. More precisely, in order to deal with zeros among the complaint variables, 0.5 times the smallest non-zero value is added before taking the logarithm.

<sup>6</sup>*Better* here means: higher satisfaction, more agreement with the list of beneficiaries, fewer instances of poor households not on the list or non-poor households on the list, the targeting method being more correct and comparing favorably to the BLT method, the number of available benefits per village being closer to correct, and fewer complaints to the village head and in the complaint box.

Table 3.1.1: Outcomes from the household endline survey

Treatment group	PMT	Hybrid	Community
Are you satisfied with the targeting activities in this village in general? (1 = worst, 4 = best)	3.042 (0.0367)	3.080 (0.0367)	3.280 (0.0326)
Do you agree with the households on the list of targeted households? (0 = no, 1 = yes)	0.669 (0.0221)	0.778 (0.0192)	0.878 (0.0149)
Are there any poor households that should be added to the list? (0 = no, 1 = yes)	0.579 (0.0231)	0.600 (0.0225)	0.389 (0.0223)
Are there any non-poor households that should be subtracted from the list? (0 = no, 1 = yes)	0.456 (0.0233)	0.357 (0.0220)	0.220 (0.0189)
Is the method applied to determine the targeted households appropriate? (1 = worst, 4 = best)	3.243 (0.0400)	3.228 (0.0418)	3.388 (0.0348)
How does this method compare to other methods (like BLT) in targeting households? (1 = worse, 3 = better)	2.546 (0.0337)	2.520 (0.0362)	2.615 (0.0315)
Is the number of households on the list too small, correct, or too large? (1 = too small, 2 = correct, 3 = too large)	1.611 (0.0283)	1.595 (0.0286)	1.716 (0.0268)
Number of interviewed households	465	480	490

The table shows group means, with standard errors in parentheses.

Table 3.1.2: Outcomes from the village head endline survey and the complaints box

Treatment group	PMT	Hybrid	Community
In your opinion, are villagers satisfied with the targeting activities in this village in general? (1 = worst, 4 = best)	2.456 (0.0610)	2.986 (0.0525)	3.389 (0.0451)
Are there any poor households that should be added to the list? (0 = no, 1 = yes)	0.732 (0.0307)	0.673 (0.0319)	0.565 (0.0340)
Are there any non-poor households that should be subtracted from the list? (0 = no, 1 = yes)	0.0574 (0.0161)	0.0369 (0.0128)	0.0467 (0.0145)
Is the method applied to determine the targeted households appropriate? (0 = no, 1 = yes)	0.565 (0.0344)	0.753 (0.0295)	0.939 (0.0165)
How does this method compare to other methods (like BLT) in targeting households? (1 = worse, 3 = better)	2.236 (0.0561)	2.610 (0.0448)	2.821 (0.0318)
How many households complained about the list of targeted households? (per household)	0.0849 (0.00947)	0.0539 (0.00621)	0.0337 (0.00529)
Number of complaints in the complaint box (per household)	0.0392 (0.00748)	0.0287 (0.00637)	0.0145 (0.00258)
Is the number of households on the list too small, correct, or too large? (1 = too small, 2 = correct, 3 = too large)	1.309 (0.0404)	1.358 (0.0377)	1.466 (0.0408)
Number of villages	209	217	214

The table shows group means, with standard errors in parentheses.

## 3.2 Empirical Model

In this section, I first introduce a baseline model of satisfaction. Two of its components—targeting accuracy and local poverty—depend on the choice of a welfare measure. The construction these components and of the different welfare measures is discussed in the subsections further below.

### 3.2.1 Model of Satisfaction

I propose a linear regression model to explain satisfaction:

$$y_{ij} = \beta_0 + \beta_1 I_H + \beta_2 I_C + \beta_3 t_j^m + \beta_4 b_j + \beta_5 h_j^m + \beta_6 x_{ij} + \varepsilon_{ij}. \quad (3.2.1)$$

$y_{ij}$  stands for any of the satisfaction outcomes in Tables 3.1.1 and 3.1.2 for household  $i$  in village  $j$ . For village level outcomes, subscript  $i$  becomes obsolete.  $I_H$  and  $I_C$  are indicators for the Hybrid and the Community treatment group, respectively, with the PMT group being the excluded category.  $t_j^m$  refers to within-village targeting accuracy, i.e., the share of correctly targeted households given the available benefits, based on welfare measure  $m$ .  $b_j$  denotes the benefit ratio, i.e., the number of benefits over the number of households  $n_j$ .  $h_j^m$  denotes the local poverty headcount ratio, which depends on the underlying welfare measure  $m$ .  $x_{ij}$  is a vector of household and village characteristics that may potentially affect satisfaction as well as targeting accuracy or benefit ratio. These characteristics include regional dummies (for all combinations of the three provinces and urban/rural), log village size, and—only for the outcomes from the endline household survey—dummies for whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of the two.

Coefficients  $\beta_1$  and  $\beta_2$  indicate what difference the treatment group makes for satisfaction, after controlling for how well the different treatments work at distributing funds to the poor. In other words, they should tell how much participation is intrinsically valued.  $\beta_3$  and  $\beta_4$  reveal the relative importance of (mis)allocation within villages and the total amounts to be distributed to each village. Lastly, using different measures  $m$  can show how well different concepts of welfare are able to explain the variation in satisfaction.

### 3.2.2 Within-village Targeting Accuracy and Welfare Measures

Within-village targeting accuracy shows how well the benefits available for a given village are distributed. Let  $b_{ij}$  be an indicator that household  $i$  in village  $j$  with  $n_j$  households received the benefit, and let  $p_{ij}^m$  be an indicator that household  $i$  is among the poorest

$\sum_{k=1}^{n_j} b_{kj}$  households in village  $j$  according to welfare measure  $m$ . Within-village targeting accuracy<sup>7</sup> is defined as

$$t_j^m := n_j^{-1} \sum_{k=1}^{n_j} (b_{kj} p_{kj}^m + (1 - b_{kj}) (1 - p_{kj}^m)). \quad (3.2.2)$$

While the recipient households are fixed, the within-village rankings of welfare—and thus  $p_{ij}^m$ —depend on how welfare  $m$  is being defined. (In what follows, superscripts  $m$  are omitted where it serves readability.) One way to do so implicitly is to assume that welfare is perfectly observed and reported by the participants of the PWR, which is sufficient to define  $t_j$ . I call this approach *rank-consistent welfare*. It implies that for the Community treatment group, targeting accuracy equals 1 for each village, as benefits were given to the  $\sum_k b_{kj}$  lowest ranked households. For the Hybrid treatment group, targeting accuracy would be lower than 1 on average, as the allocation among the lowest ranked  $1.5 \cdot n_j b_j$  households was determined by a PMT score ranking instead of the PWR. For the PMT group, targeting accuracy is not observed, as the villages in this group did not conduct a PWR.

An alternative way to compute targeting accuracy using PWRs, which does allow to include households that were not ranked, is to use predicted values based on a latent welfare model with household characteristics. I call this approach *rank-score welfare* (the word *score* indicating that this is an estimated rather than observed ranking). The predicted ranking outcomes will inevitably differ to some degree from the order of the actual PWRs. This may be seen as a disadvantage, as the true PWRs may reflect some important factors that are not being asked for in the surveys, for instance because they are sensitive or hard to quantify. But it may also be seen as an advantage, as the predicted rankings only take into account the factors that are considered relevant in multiple villages, while idiosyncratic factors such as a household's popularity or connectedness to local elites are left out of the score—a desirable effect.

Lastly, per capita consumption also constitutes a welfare measure. The BPS data does not contain detailed enough information to construct these figures, though, so that only predicted values can be used here as well.<sup>8</sup> To train the consumption model, I use the

---

<sup>7</sup>There are other measures of targeting quality than accuracy, that assign different weights to poor non-beneficiaries (type 1 errors) and non-poor beneficiaries (type 2 errors), or factor in the severity of household poverty (see e.g. Ravallion, 2009). These are, however, largely incompatible with the ordinal nature of a rank-based welfare measure, and with the fact that for within-village targeting, any type 1 error automatically also results in a type 2 error.

<sup>8</sup>Importantly, the PMT score, that also aimed at predicting per capita consumption based on the BPS data, does a much poorer job than the prediction model used in this chapter: for the PMT group, the correlation coefficient of log per capita consumption with the log PMT score is 0.53, while the correlation



consumption figures from the baseline survey. In the following paragraphs, I outline how consumption and rank-score welfare are being constructed.

To estimate the relation between welfare  $w_{ij}$  of household  $i$  in village  $j$  and observable household characteristics  $z_{ij}$ , I assume a linear model,

$$w_{ij} = z_{ij}\gamma + \xi_{ij}. \quad (3.2.3)$$

When the welfare measure to be predicted is (log) consumption, 3.2.3 can be estimated via OLS. For rank-score welfare, I propose to use a rank-ordered logit (ROL) model (Beggs, Cardell and Hausman, 1981), which was originally developed to study consumer preferences. The key assumption that needs to be made is that the random disturbance term  $\xi_{ij}$  in 3.2.3 is iid type I extreme value (EV1) distributed. In accordance with levels of welfare, village  $j$  provides a complete ranking over the set of households,  $R_j$ . For ease of notation, assume  $w_{1j} < \dots < w_{n_jj}$ . The zero-probability case of equal welfare of two households is being ignored here. The probability for any particular ranking to occur given household characteristics  $Z_j = (z_{1j}, \dots, z_{n_jj})'$  is then

$$\Pr[R_j|Z_j; \gamma] = \Pr[w_{1j} < \dots < w_{n_jj}|Z_j; \gamma] \quad (3.2.4)$$

$$= \Pr[w_{1j} < w_{2j}|Z_j; \gamma] \cdot \Pr[w_{ij} < w_{3j} \forall i = 1, 2|Z_j; \gamma] \cdot \dots \\ \dots \cdot \Pr[w_{ij} < w_{n_jj} \forall i = 1, \dots, n_j - 1|Z_j; \gamma] \quad (3.2.5)$$

$$= \prod_{k=2}^{n_j} \frac{\exp(z_{kj}\gamma)}{\sum_{l=1}^k \exp(z_{lj}\gamma)}. \quad (3.2.6)$$

The ROL formula is a product of multinomial logit probabilities. Step 3.2.5 follows from the assumption that the conditional distribution of the highest ranked household from any subset is independent of the ranking of the other households. This is equivalent to the irrelevance of independent alternatives (IIA) property, which follows from the EV1 specification of  $\xi_{ij}$ . The reduction to the above closed-form expression allows estimation via maximum likelihood. The log-likelihood for the sample is

$$L(\gamma) = \sum_{j=1}^J \ln(\Pr[R_j|Z_j; \gamma]). \quad (3.2.7)$$

Welfare scores  $\hat{w}_{ij}$  can be then be predicted using estimated coefficients  $\hat{\gamma}$  just the same way as in the model of consumption,  $\hat{w}_{ij} = z_{ij}\hat{\gamma}$ . These scores are ordinal, i.e., unlike

---

coefficient with the predicted per capita consumption estimated here is 0.72. This leads to large variation in within-village targeting accuracy (based on predicted consumption), which is needed to identify its impact on satisfaction.

consumption scores they are not interpretable on their own.

### 3.2.3 Local Poverty Rate

An important control variable is the poverty headcount for each village. Constructing this requires—in addition to welfare scores that allow comparisons of households between villages—a global poverty threshold. While for some welfare measures, such as consumption, there are natural poverty thresholds such as the 2\$ per day poverty line, there is none such for the rank-score welfare measure. For this chapter, I set the overall poverty rate equal to the overall benefit ratio. This has the advantage that the total poverty rate is the same across different welfare measures. It is also consistent with the poverty rate implied by the 2\$ per day consumption poverty line, which the total benefit ratio in the field experiment aimed to meet.

Village poverty rates are computed as follows: welfare scores  $\hat{w}_{ij}^m$  of households in the entire sample are being ranked. The lowest ranked households are declared poor,  $h_{ij}^m = 1$ , and the remaining ones non-poor,  $h_{ij}^m = 0$ , with the threshold being the total number of benefits allocated,  $\sum_l \sum_k b_{kl}$ . The local poverty rate of village  $j$  is then defined as

$$h_j^m := n_j^{-1} \sum_{k=1}^{n_j} h_{kj}^m. \quad (3.2.8)$$

### 3.2.4 Total Targeting Accuracy

Total targeting accuracy  $T_j^m$  is constructed just like within-village targeting accuracy, but using the global poverty indicators  $h_{ij}^m$ ,

$$T_j^m := n_j^{-1} \sum_{k=1}^{n_j} (b_{kj} h_{kj}^m + (1 - b_{kj}) (1 - h_{kj}^m)). \quad (3.2.9)$$

There are a number of reasons for using within-village targeting accuracy instead of total targeting accuracy in model 3.2.1. To begin with, it is conceivable that mistargeting within and between villages is perceived differently. Given that the targeting process has two stages, targeting errors have their origins within and outside the village, which may affect satisfaction in different ways. Violations of the ordering within the village may be perceived stronger and as more unfair than missing the correct number of benefits given the regional poverty line. In addition, for some of the welfare models being used it is impossible to construct total targeting accuracy, while for others it is possible but at the cost of relatively lower precision. Despite these caveats, total targeting accuracy may

be a meaningful metric, and it is being used as a robustness check when comparing the explanatory power of different welfare models.

### 3.3 Model Selection and Estimation

In this section, I first line out how the welfare scores are being constructed given a set of predictors. I then go on to describe how the model of satisfaction is being estimated given certain limitations in the data.

#### 3.3.1 Model Selection

To account for regional differences in the factors and weights constituting welfare, each welfare model is being estimated separately by province (North Sumatra, South Sulawesi, and Central Java) and urban/rural areas. In order to achieve high predictive performance, for each model and each sample, I use an alternating forward/backward model selection procedure in order to identify an adequate set of variables from a list of potential predictors. The procedure minimizes the bias-corrected Akaike information criterion, as defined by Hurvich and Tsai (1989), which is asymptotically equivalent to leave-one-out cross-validation. The candidate variables are taken from Table 3.A.1, extended by polynomial terms and logarithms of age, household size, and floor area per capita, as well as by interaction terms of gender and marriage status, of age and education, of education and occupational sector, as well as a number of cluster-level variables and interactions of all individual-level and cluster-level variables.

There are some subtle differences in the respective variable pools for the different models. For the models of consumption, including village- or higher level variables as predictors may help to increase the model fit. On the other hand, this is pointless when using the ranking outcomes, as  $\gamma$  is identified only from comparisons of households within villages. The inability of the ROL model to incorporate village-level variation is not relevant for the construction of  $t_j^{\text{rank-score}}$ , but it may render  $h_j^{\text{rank-score}}$  less reliable when compared to the version based on consumption,  $h_j^{\text{consumption}}$ . To what extent this is true depends on how much of the variation in welfare is captured by household-level variables relative to village-level variables. For the construction of within-village targeting accuracy  $t_j^{\text{rank-score}}$ , it may be helpful to include interactions of household-level variables with village-level variables. These can emphasize local differences in the relative importance of certain household level factors for poverty, and thus improve the estimated rankings. On the other hand, for measures based on global rank-score poverty indicators, such as  $h_j^{\text{rank-score}}$  and  $T_j^{\text{rank-score}}$ , interactions should not be included, as they would shift and scale welfare scores

differently for each village. Since the ROL model does not actually relate households from different villages to each other, such transformations would be arbitrary, and render comparisons between villages impossible. This means, in particular, that total targeting accuracy  $T_j^{\text{rank-score}}$  is constructed with a smaller set of potential predictors than within-village targeting accuracy  $t_j^{\text{rank-score}}$ , presumably leading to lower precision of the former compared to the latter.

The prediction models of consumption are estimated using all the households in the baseline survey. The prediction models of rank-score welfare are estimated based on the BPS data of the Hybrid group. To prevent overfitting due to using the same villages both for estimation and prediction, I use a one-village-out cross-fitting procedure: for the predicted ranking in village  $j$ , the estimation includes all villages except  $j$ .

To construct  $t_j$  and  $h_j$  with a predicted ranking as benchmark, the joint distribution of  $(b_{kj}, z_{kj})_{k=1, \dots, n_j}$  is required.  $b_{ij}$  is known for every household in the PMT and the Hybrid group, but  $z_{ij}$  is only fully observed for the PMT group, while in the Hybrid group it is observed only for the  $1.5 \cdot n_j b_j$  lowest-ranked households from the PWR. Therefore, for estimations that involve the Hybrid group, I use imputed values  $\hat{t}_j$  and  $\hat{h}_j$ . The imputation procedure is outlined in section 3.A.1 of the appendix, together with an assessment of the bias arising from it. Table 3.3.1 gives a summary of all the information available for each treatment group.

### 3.3.2 Estimation

Due to the fact that ranking and consumption data is not available for every household of the experiment, equation (3.2.1) cannot be estimated directly. In particular, it is not possible to construct targeting accuracy and local poverty based on rank-score welfare or predicted consumption for the villages in the Community group,  $J_C$ , since the kind of data collected by the BPS for every household to construct PMT scores was not collected there. On the other hand, for the Hybrid group  $J_H$  and the Community group  $J_H$ , it is possible to construct targeting accuracy based on the true rankings—but not for the PMT group  $J_P$ , as no PWR was conducted there. To get around this issue, I conduct separate sets of estimations.

The first one uses rank-score welfare and only villages from the Hybrid group and the PMT group,

$$y_{ij} = \beta_0 + \beta_1 I_H + \beta_3 \hat{t}_j^{\text{rank-score}} + \beta_4 b_j + \beta_5 \hat{h}_j^{\text{rank-score}} + \beta_6 x_{ij} + \varepsilon_{ij}, \quad j \in J_P \cup J_H. \quad (3.3.1)$$

Table 3.3.1: Overview of data sources and constructed variables by treatment group

	PMT group	Hybrid group	Community group
Baseline survey (contains ranks, consumption figures, detailed household characteristics)	9 households from each village	9 households from each village only lowest ranked households (~60%)	9 households from each village
BPS data (contains ranks, some household characteristics and assets)	all households		-
Endline survey for village leaders (contains satisfaction data)	all villages	all villages	all villages
Endline survey for households (contains satisfaction data)	3 households, only Central Java villages	3 households, only Central Java villages	3 households, only Central Java villages
$t_j^{\text{rank-score}}, T_j^{\text{rank-score}}, h_j^{\text{rank-score}}$	x	-	-
$\hat{t}_j^{\text{rank-score}}, \hat{T}_j^{\text{rank-score}}, \hat{h}_j^{\text{rank-score}}$	x	x	-
$t_j^{\text{consumption}}, T_j^{\text{consumption}}, h_j^{\text{consumption}}$	x	-	-
$\hat{t}_j^{\text{consumption}}, \hat{T}_j^{\text{consumption}}, \hat{h}_j^{\text{consumption}}$	x	x	-
$t_j^{\text{rank-consistent}}, T_j^{\text{rank-consistent}}$	-	x	x
$b_j$	x	x	x

$t_j^m$ ,  $T_j^m$ , and  $h_j^m$ , respectively, denote within-village targeting accuracy, total targeting accuracy, and local poverty of village  $j$  and according to welfare measure  $m$ . The hat notation marks imputed variables, based only on the poorest households from each village.  $b_j$  denotes the benefit ratio. “x” indicates that the measure can be constructed for the respective treatment group.

The PMT group is the excluded category. Estimating equation 3.3.1 shows the difference in intrinsic value between the PMT and the Hybrid ranking method, as well as the relative significance of targeting accuracy.

The next set of estimations uses within-village targeting accuracy based on rank-consistent welfare, and can only be applied to the Hybrid group and the Community group:

$$y_{ij} = \beta_0 + \beta_2 I_C + \beta_3 t_j^{\text{rank-consistent}} + \beta_4 b_j + \beta_5 a_j + \beta_6 x_{ij} + \varepsilon_{ij}, \quad j \in J_H \cup J_C. \quad (3.3.2)$$

The Hybrid group is the excluded category. For the villages of the Community group, local poverty rates are not available. Instead, I include the village attendance rate for the PWR,  $a_j$ , as an additional control. Since the ranking procedure and the allocation mechanism were only explained at the meetings,  $a_j$  should be independent of treatment status, but it may proxy affluence and social capital, which could affect both targeting accuracy as well as satisfaction. Estimating equation 3.3.2 shows the difference in intrinsic value between the Hybrid and the Community ranking method.

A further set of estimations has the objective to compare the ability of rank-score welfare and predicted consumption to explain satisfaction. This is done by pooling the different measures of targeting accuracy and local poverty in one equation:

$$y_{ij} = \beta_0 + \beta_{31} t_j^{\text{rank-score}} + \beta_{32} t_j^{\text{consumption}} + \beta_4 b_j + \beta_{51} h_j^{\text{rank-score}} + \beta_{52} h_j^{\text{consumption}} + \beta_6 x_{ij} + \varepsilon_{ij}, \quad j \in J_P. \quad (3.3.3)$$

A significant estimate of, say, coefficient  $\beta_{31}$  means that targeting accuracy based on the rank-score likely helps to explain the respective satisfaction outcome, given the information provided by the targeting accuracy measure that is based on consumption. Equation 3.3.3 focuses on the PMT group, to avoid having to use imputations for targeting accuracy or local poverty. Leaving out households from the Hybrid group also rules out a possible bias caused by the degree to which the final allocation resembles the PWRs. For instance, if people in the Hybrid group are asked to rank households, but then those rankings are not being adhered to due to the second-stage ranking, people may feel actively ignored, leading to lower satisfaction that would falsely be attributed to targeting accuracy alone. These issues also apply to estimating 3.3.1, and will be addressed in the next section.

Equations 3.3.1 to 3.3.3 are estimated via OLS.<sup>9</sup> For outcomes measured at the vil-

---

<sup>9</sup>Some of the outcomes  $y_{ij}$  are binary or ordered categorical. I also ran the corresponding regressions as logit and ordered logit models, respectively. The relative coefficient sizes and  $p$ -values were almost

lage level, heteroscedasticity-consistent standard errors are used. For outcomes at the household level, standard errors are clustered at the village level.

### 3.4 Results and Discussion

The results of estimating equation 3.3.1 for household level satisfaction outcomes are shown in Table 3.4.1. The table also includes the results of a simple regression of the satisfaction outcome on the treatment group, to see how much treatment effects change after including covariates. Within-village targeting accuracy appears to be important and is significant for most outcomes. Higher targeting accuracy increases satisfaction and agreement with the list of beneficiaries, reduces the chance of households not being on the list or wrong households to be included, and increases the likelihood that participants find the allocation method to be correct. The effect of the Hybrid group, on the other hand, is mostly small and insignificant, and does not show a consistent direction. However, when compared to the simple regression results, one can see that including covariates pulls the effect of the Hybrid group away from the expected direction for almost all outcomes. The effect of the benefit ratio is mostly consistent with the effect of targeting accuracy: more coverage increases satisfaction and reduces the chance of poor households being excluded. While mostly having the same sign as the effect of targeting accuracy, the magnitude of the effect of the benefit ratio is consistently lower. This means that on average, adding a benefit to a village only increases satisfaction if the recipient household is targeted correctly. Lastly, as one would expect, the impression of whether there were enough benefits is strongly affected by the benefit ratio but not by within-village targeting accuracy.

The results of the regressions of village level outcomes, reported in Table 3.4.2, show a slightly different picture. Village heads believed on average that households in the Hybrid group were more satisfied than in the PMT group, and that the targeting method was better, even after controlling for targeting accuracy. However, this notion is not supported by significant effects of the treatment group on list errors, or complaints to the village head or in the complaint box. Targeting accuracy, on the other hand, did not only increase perceived household satisfaction and correctness of the method, but also led to significantly fewer complaints. Lastly, a higher share of beneficiaries led to a higher share of village heads thinking that households were satisfied, fewer excluded poor households, more non-poor households receiving benefits, and fewer complaints.

As mentioned earlier, two factors could bias the results in Tables 3.4.1 and 3.4.2: tar-  
indistinguishable from those of the OLS results.

Table 3.4.1: Simple treatment effects and estimation results of equation 3.3.1 – household level outcomes

	Satisfied	Agree with list	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Enough benefits
Hybrid group (simple regression)	0.0384 (0.0747)	0.110*** (0.0402)	0.0202 (0.0459)	-0.0992** (0.044)	-0.0147 (0.0779)	-0.0260 (0.0727)	-0.0146 (0.0548)
Adj. $R^2$	-0.001	0.014	-0.001	0.009	-0.001	-0.001	-0.001
Hybrid group	-0.0225 (0.0727)	0.0777* (0.042)	0.0548 (0.047)	-0.0507 (0.0443)	-0.0555 (0.0778)	-0.0201 (0.0773)	-0.0368 (0.0557)
Targeting accuracy	1.164*** (0.447)	0.510** (0.248)	-0.730** (0.324)	-0.974*** (0.271)	1.009** (0.510)	0.0681 (0.424)	0.261 (0.371)
Benefit ratio	0.515* (0.281)	0.200 (0.138)	-0.293* (0.166)	0.0383 (0.151)	0.166 (0.305)	-0.0985 (0.248)	0.614*** (0.190)
Adj. $R^2$	0.068	0.074	0.028	0.056	0.042	0.021	0.056
Observations	807	924	930	934	728	753	938

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are clustered at the village level. Households are from the PMT or Hybrid group. The first rows show the results of a simple regression of the outcome on the Hybrid group dummy. Targeting accuracy and local poverty use rank-score welfare as benchmark. The multiple regression specifications control for local poverty, regional dummies, log village size, and dummies indicating whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of those two. The coefficients of these controls are omitted.



Table 3.4.2: Simple treatment effects and estimation results of equation 3.3.1 – village level outcomes

	Households satisfied	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Complaints to village head (log)	Complaints in complaint box (log)	Enough benefits
Hybrid group (simple regression)	0.530*** (0.0805)	-0.0592 (0.0443)	-0.0205 (0.0206)	0.189*** (0.0453)	0.375*** (0.0718)	-0.317** (0.140)	-0.198 (0.125)	0.0493 (0.0553)
Adj. $R^2$	0.092	0.002	0.000	0.037	0.059	0.010	0.004	0.000
Hybrid group	0.435*** (0.0885)	-0.0504 (0.0448)	-0.0106 (0.0224)	0.143*** (0.0484)	0.325*** (0.0785)	-0.194 (0.147)	-0.0992 (0.108)	0.0531 (0.0559)
Targeting accuracy	1.502*** (0.563)	-0.233 (0.337)	-0.184 (0.187)	0.912*** (0.311)	0.86 (0.540)	-2.201** (0.993)	-1.330* (0.688)	0.00834 (0.443)
Benefit ratio	0.823*** (0.311)	-0.887*** (0.167)	0.251** (0.113)	0.187 (0.187)	0.173 (0.300)	-1.502*** (0.543)	-1.082** (0.438)	1.099*** (0.234)
Adj. $R^2$	0.127	0.094	0.092	0.069	0.062	0.080	0.378	0.136
Observations	421	426	426	424	421	426	426	419

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
Estimations are done by OLS. Standard errors (in parentheses) are heteroscedasticity-consistent. Households are from the PMT or Hybrid group. The first row shows the results of a simple regression of the outcome on the Hybrid group dummy. Targeting accuracy and local poverty use rank-score welfare as benchmark. The multiple regression specifications control for local poverty, regional dummies, and log village size. The coefficients of these controls are omitted.

getting accuracy and local poverty are imputed, and the targeting measure and allocation method more or less coincide for the Hybrid group. Table 3.A.2 in the appendix shows the results of the same estimations for the PMT group only, and using true targeting accuracy and local poverty. The coefficients of targeting accuracy and benefit ratio are statistically indistinguishable from those in Tables 3.4.1 and 3.4.2 and show no evidence of bias in any direction.

Tables 3.4.3 and 3.4.4 show the results of estimating equation 3.3.2. The household level outcomes show fewer significant impacts of targeting accuracy (which is now using actual instead of predicted rankings as benchmark). This is likely due to higher standard errors caused by the lack of variation in targeting accuracy for the Community treatment group. However, the magnitudes in the effects on satisfaction and indicators of correctness of the beneficiary list are comparable to those in Table 3.4.1. The village level outcomes strongly suggest that targeting accuracy increases household satisfaction and correctness of list and method, and reduces complaints to the village head. At the same time, the effect of the Community group (versus the Hybrid group) is mostly insignificant and fairly inconsistent in sign. This can partly be attributed to large standard errors, stemming from a high correlation between targeting accuracy and treatment group. When comparing the coefficients between the simple and multiple regressions, it becomes evident that including targeting accuracy and other covariates strongly pulls the effect of the treatment group away from its expected direction. Lastly, a higher share of beneficiaries increased satisfaction and perceived correctness of list and method, and reduced complaints. Again, the magnitude of these effects appears lower than that of within-village targeting accuracy.

In summary, household members and village heads seem to notice and appreciate increases in targeting accuracy and the number of benefits. The effects of the treatment group on satisfaction-related outcomes are mostly small and statistically insignificant. The exception is that village heads perceived households to be more satisfied in the Hybrid group than in the PMT group, and generally believed the former method was better than the latter. This notion, however, is not backed up by the respective counterpart variables at the household level, nor by significant effects of more tangible indicators of dissatisfaction such as complaints received. It is thus possible that, while the targeting method itself has little intrinsic value to most people, village heads have a preference for an allocation by ranking. This might be due to a sense of importance derived from being involved in the execution of the PWRs, or because PWRs were perceived to be more different than the PMT method from the allocation method of the BLT, which was deemed highly inadequate.

Tables 3.4.5 and 3.4.6 show the results of estimating equation 3.3.3. On the household

Table 3.4.3: Simple treatment effects and estimation results of equation 3.3.2 – household level outcomes

	Satisfied	Agree with list	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Enough benefits
Community group (simple regression)	0.200*** (0.069)	0.0989*** (0.032)	-0.210*** (0.045)	-0.137*** (0.040)	0.160** (0.075)	0.094 (0.067)	0.120** (0.054)
Adj. $R^2$	0.018	0.016	0.043	0.022	0.010	0.004	0.009
Community group	0.0443 (0.125)	0.0792 (0.0514)	-0.113 (0.093)	-0.00511 (0.0700)	0.154 (0.147)	-0.0137 (0.112)	0.136 (0.0939)
Targeting accuracy	0.931 (0.583)	0.0957 (0.250)	-0.54 (0.488)	-0.779** (0.355)	0.147 (0.646)	0.739 (0.525)	-0.159 (0.462)
Benefit ratio	0.516** (0.227)	0.0465 (0.0887)	-0.185 (0.146)	0.268** (0.122)	0.117 (0.264)	0.255 (0.201)	0.545*** (0.181)
Adj. $R^2$	0.094	0.057	0.075	0.060	0.105	0.054	0.085
Observations	831	955	952	958	747	767	967

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are clustered at the village level. Households are from the Hybrid or Community group. The first row shows the results of a simple regression of the outcome on the Community group dummy. Targeting accuracy uses rank-consistent welfare as benchmark. The multiple regression specifications control for attendance rates of the PWR, regional dummies, log village size, and dummies indicating whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of those two. The coefficients of these controls are omitted.

Table 3.4.4: Simple treatment effects and estimation results of equation 3.3.2 – village level outcomes

	Households satisfied	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Complaints to village head (log)	Complaints in complaint box (log)	Enough benefits
Community group (simple regression)	0.403*** (0.069)	-0.107** (0.047)	0.010 (0.019)	0.185*** (0.034)	0.210*** (0.055)	-0.410*** (0.125)	-0.116 (0.108)	0.108* (0.056)
Adj. $R^2$	0.072	0.010	-0.002	0.063	0.031	0.022	0.000	0.007
Community group	0.0544 (0.126)	0.0846 (0.0839)	0.0715** (0.0282)	0.0669 (0.0573)	-0.0451 (0.0933)	0.113 (0.214)	-0.16 (0.186)	-0.0035 (0.0946)
Targeting accuracy	2.152*** (0.668)	-1.129*** (0.421)	-0.385** (0.195)	0.733** (0.332)	1.523*** (0.556)	-3.059*** (1.132)	0.208 (0.944)	0.671 (0.519)
Benefit ratio	1.038*** (0.200)	-0.811*** (0.149)	0.0291 (0.0544)	0.059 (0.104)	0.359** (0.141)	-1.340*** (0.354)	-0.278 (0.336)	0.781*** (0.184)
Adj. $R^2$	0.137	0.081	0.053	0.072	0.055	0.097	0.294	0.076
Observations	423	431	431	427	425	431	431	423

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are heteroscedasticity-consistent. Households are from the Hybrid or Community group. The first row shows the results of a simple regression of the outcome on the Community group dummy. Targeting accuracy uses rank-consistent welfare as benchmark. The multiple regression specifications control for attendance rates of the PWR, regional dummies, and log village size. The coefficients of these controls are omitted.

level, the results indicate that targeting accuracy based on rank-score welfare is significantly predictive for most satisfaction outcomes—even conditional on targeting accuracy measured by predicted consumption. The reverse is not true for any outcome. On the village level, the same conclusion holds for the number of complaints to the village head and in the complaint box. Taken together, these results suggest that targeting accuracy explains satisfaction better when it is based on predicted ranks rather than predicted consumption.

One reason for this might be that the consumption model simply has a lower predictive performance than the model of ranks. After all, the two models are estimated based on different samples: the consumption model uses the baseline survey of 9 households per village, whereas the rank model uses all the households from the Hybrid group in the BPS data. To rule out this potential element of unfairness in the comparison, I create an alternative set of welfare scores based on a ROL model that uses the baseline survey instead. Tables 3.A.3 and 3.A.4 in the appendix show estimations of equation 3.3.3 with targeting accuracy and local poverty based on these new welfare scores. The results remain qualitatively the same. Tables 3.A.5 and 3.A.6 in the appendix display estimation results of estimating equation 3.3.3 using total targeting accuracy (and therefore leaving out the benefit ratio). The results confirm the above findings, and show significant effects of targeting accuracy based on rank-score welfare even for most of the village level outcomes. Lastly, to confirm that the results are not a coincidental product of focusing on the PMT group only, I also run the same regressions including the Hybrid group, using imputed targeting accuracy and local poverty and including a treatment group dummy. The results, reported in Tables 3.A.7 and 3.A.8 in the appendix, again remain qualitatively unchanged.

The way in which local poverty would impact program satisfaction is not clear a priori. However, one would expect that, controlling for the benefit ratio, villages with higher poverty incidence register more omitted poor households and fewer targeted non-poor households, and a comparatively less sufficient amount of benefits. The results in the tables above and the appendix confirm this pattern for local poverty measured against rank-score welfare, but not for consumption. It is noteworthy that the poverty rate based on the ROL model—which does not take into account village level predictors and does not compare households between villages—is able to pick up these relationships better than the poverty rate based on a consumption regression model that does include village level predictors.

Table 3.4.5: Estimation results of equation 3.3.3 – household level outcomes

	Satisfied	Agree with list	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Enough benefits
Targeting accuracy	1.278***	0.811**	-0.31	-1.212***	1.125**	0.438	0.917**
(rank-score welfare)	(0.473)	(0.334)	(0.377)	(0.323)	(0.552)	(0.504)	(0.427)
Targeting accuracy	-0.649	0.062	-0.0641	-0.23	-0.596	-1.120*	-0.737
(predicted cons.)	(0.695)	(0.384)	(0.421)	(0.383)	(0.684)	(0.643)	(0.518)
Local poverty rate	-0.135	0.316*	0.409*	-0.489**	-0.162	0.0475	-0.463*
(rank-score welfare)	(0.252)	(0.180)	(0.238)	(0.197)	(0.337)	(0.330)	(0.266)
Local poverty rate	-0.420*	-0.205	-0.0298	0.411**	-0.219	-0.598***	-0.0183
(predicted cons.)	(0.231)	(0.172)	(0.166)	(0.172)	(0.290)	(0.202)	(0.201)
Benefit ratio	0.406	0.282*	-0.0383	-0.243	-0.175	-0.0763	0.244
	(0.337)	(0.154)	(0.216)	(0.176)	(0.389)	(0.292)	(0.236)
Adj. $R^2$	0.088	0.080	0.012	0.088	0.041	0.024	0.062
Observations	383	453	456	458	342	357	460

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
 Estimations are done by OLS. Standard errors (in parentheses) are clustered at the village level. Households are from the PMT group only. All specifications control for regional dummies, log village size, and dummies indicating whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of those two. The coefficients of these controls are omitted.

Table 3.4.6: Estimation results of equation 3.3.3 – village level outcomes

	Households satisfied	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Complaints to village head (log)	Complaints in complaint box (log)	Enough benefits
Targeting accuracy (rank-score welfare)	0.744 (0.603)	0.0456 (0.330)	-0.0729 (0.188)	0.467 (0.393)	1.019 (0.619)	-4.341*** (1.057)	-1.720** (0.689)	-0.146 (0.435)
Targeting accuracy (predicted cons.)	1.14 (1.020)	-0.034 (0.402)	0.16 (0.205)	0.348 (0.517)	-0.43 (0.883)	1.876 (1.407)	1.275 (1.113)	0.0421 (0.527)
Local poverty rate (rank-score welfare)	0.474 (0.451)	0.344 (0.232)	-0.223* (0.135)	0.468* (0.246)	0.700* (0.405)	-1.240* (0.684)	0.252 (0.640)	-0.644** (0.324)
Local poverty rate (predicted cons.)	-0.0593 (0.371)	0.207 (0.184)	0.028 (0.121)	-0.163 (0.207)	-0.0872 (0.354)	0.952 (0.581)	0.314 (0.466)	-0.171 (0.269)
Benefit ratio	1.292*** (0.483)	-0.944*** (0.228)	0.336** (0.166)	0.256 (0.261)	0.166 (0.434)	-1.499** (0.750)	-1.051* (0.612)	1.181*** (0.347)
Adj. $R^2$	0.027	0.114	0.066	0.003	-0.023	0.136	0.435	0.126
Observations	206	209	209	209	208	209	209	204

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are heteroscedasticity-consistent. Households are from the PMT group only. All specifications control for regional dummies and log village size. The coefficients of these controls are omitted.

### 3.5 Conclusion

The findings in this chapter suggest that PWRs are popular largely due to their outcome, not the procedure. Furthermore, allocations coinciding with scores from PWRs were better-received than those coinciding with consumption scores. These insights could lead to improvements in targeting: similar to PMT scores of consumption, one could use welfare scores, trained on PWRs for a representative sample of villages. Then, the households with the lowest scores below a predefined threshold are targeted. The targeting threshold could come from a poverty map or be set according to budgetary restrictions. It could also come from the PWR itself, for instance if it takes the form of a poverty classification exercise instead of a continuous welfare ranking. The resulting scores combine the respective strengths of proxy means tests and participatory approaches, in being objective and built on observable factors while being founded in local views of welfare.

The findings may prove particularly useful for contexts in which conducting PWRs in all target locations may not be feasible or desirable. This is the case for settings where people cannot be expected to know most of the households in their neighborhood, or where poverty is not deemed an appropriate topic to be discussed publicly. Further worries about large scale PWRs might include selfish ranking behavior, discrimination, elite capture, or coercion. Using welfare scores based on training PWRs without rank-dependent payouts should mitigate such concerns.

Constructing and applying welfare scores based on PWRs leads to further interesting problems and opens up several pathways for exploration. One avenue is refining the rank prediction model. It may be worth it to take a closer look at the factors used, and to consider further determinants of the rankings—particularly those related to welfare potential rather than current material well-being—to be able to craft more focused questionnaires and eventually more precise welfare scores. At the same time, the best fitted model may not always be the most preferable one. The approach in this chapter has been to treat PWR results as correct representations of welfare. Of course, this is adequate only if the rankings contain no systematic biases on grounds of ethnicity, religion, gender, etc. Going forward, it will be useful to watch out for such biases and understand the local conditions that lead to them. Once identified, it may be prudent to purposefully leave out any markers or correlates of traits eliciting such biases. Taking this one step further, one could apply statistical debiasing techniques to the prediction algorithm, and thus weed out biases against specific traits completely. I regard exploring such improvements of participatory welfare scores, and their impact on allocations and satisfaction, as a worthwhile endeavor for future research.



## 3.A Appendix

### 3.A.1 Imputation Procedure

In this section, I address the issue that the better-off households (according to the PWR) in the Hybrid group do not appear in the BPS data. I subsume the Hybrid group households represented in the BPS dataset under  $H_1^{\text{Hybrid}}$  and the other ones under  $H_2^{\text{Hybrid}}$ . If we use actual rankings as benchmark, the fact that the households in  $H_2^{\text{Hybrid}}$  are unobserved is not a problem, as they are all counted as non-poor and did not receive the benefit, and thus all count as correctly targeted. However when using predicted welfare as benchmark, it cannot be ruled out that the model would have classified some of the unobserved households as poor. Thus, in assuming that all of them were targeted correctly we would overstate targeting accuracy.

I propose to look at the households within  $H_2^{\text{Hybrid}}$  that were visited for the baseline survey,  $H_2^{\text{Hybrid, baseline}}$ , to get an average targeting accuracy for the unobserved part in each village. To do so, I go through the following steps.

1. For the PMT group, I establish which households would most likely have been included in the BPS survey, had they been in the Hybrid group instead. This is to make predictions in the PMT group just as in the Hybrid group. With the households in the BPS survey, I estimate a logit model of treatment status on household characteristics, and assign propensity scores to the households of the PMT group. The lower these are, the less likely it is to find a similar household in the Hybrid group, meaning that its equivalent in the Hybrid group would less likely have been included in the BPS data. I then rank households in each village  $j$  of the PMT group by propensity score and subsume the  $\min\{1.5 \cdot n_j b_j, n_j\}$  highest scoring households under the set  $H_1^{\text{PMT}}$ , and the remaining ones under  $H_2^{\text{PMT}}$ , with subset  $H_2^{\text{PMT, baseline}}$  of households observed in the baseline survey.
2. I predict welfare for the households in  $H_1^{\text{PMT}} \cup H_2^{\text{PMT}}$  and rank all of them according to their score. I define the absolute poverty line as the welfare score of the  $B^{\text{PMT}th}$  poorest household, with  $B^{\text{PMT}}$  the number of benefits made available to all households in the PMT group. Furthermore, for each village  $j$  in the PMT group I compute  $t_j$  and  $h_j$ .
3. I predict welfare for the households in  $H_1^{\text{Hybrid}} \cup H_2^{\text{Hybrid, baseline}}$ . Then for each village  $j$ , both in the PMT group and the Hybrid group, I compute the *reduced* targeting accuracy  $t_{1,j}$ , that pretends the households in  $H_1^{\text{PMT}} \cup H_1^{\text{Hybrid}}$  make up the entire village.

4. Using the households in  $H_2^{\text{PMT, baseline}}$  and  $H_2^{\text{Hybrid, baseline}}$ , respectively, I calculate  $t_2^{\text{PMT}}$  and  $t_2^{\text{Hybrid}}$ —the overall fractions of households below the local poverty lines used to calculate  $t_{1,j}$  in the step above. These averages are used as a proxy for mistargeting among the households unobserved by the BPS. Since the number of households per village in  $H_2^{\text{PMT, baseline}}$  and  $H_2^{\text{Hybrid, baseline}}$  is very low—0 in some cases—it is not feasible to use village-specific averages instead.
5. Eventually, the two separate estimates are combined in the following way. I construct two factors,  $z_{1,j} = s_j t_{1,j}$  and  $z_{2,j} = (1 - s_j) t_{2,g}$ , where  $s_j$  is the fraction of households from  $H_1^{\text{PMT}} \cup H_1^{\text{Hybrid}}$  in village  $j$ . I then regress  $t_j$  on  $z_{1,j}$  and  $z_{2,j}$  for the PMT villages and construct linear predictions,  $\hat{t}_j$ , for both the PMT group and the Hybrid group.

The reason for not simply adding up  $z_{1,j}$  and  $z_{2,j}$  is that both  $t_{1,j}$  and  $t_{2,g}$  are not unbiased estimates of their respective shares of targeting accuracy:  $t_{1,j}$  does not take into account that the poorer households in  $H_2^{\text{PMT}} \cup H_2^{\text{Hybrid}}$  may render some of the households declared poor in  $H_1^{\text{PMT}} \cup H_1^{\text{Hybrid}}$  non-poor, and  $t_{2,g}$  does not take into account that this is likely to change the local poverty line.

Just as targeting accuracy, the local poverty rate  $h_j$  also needs to be estimated, since the total number of poor households is not observed for the Hybrid group. I take a slightly different approach than for targeting accuracy, though, as the treatment group should have no influence on the total amount of benefits (which was determined from a government census, independently of the treatment group assignment). Therefore, predictions  $\hat{h}_j$  of the poverty headcount for the PMT and the Hybrid group are made with the PMT group as training data. As candidate predictors I use dummies for region, (log) village size, the fraction of observed households  $s_j$ , and the poverty headcount only considering the households in  $H_1^{\text{PMT}} \cup H_1^{\text{Hybrid}}$ , as well as various transformations and interactions of these variables. Just as in the welfare models, I use a stepwise model selection procedure as well as cross-fitting to prevent overfitting.

Comparing the effects of  $t_j$  and  $\hat{t}_j$  as well as  $h_j$  and  $\hat{h}_j$  for rank-score welfare in the PMT group shows that the imputation does not significantly change measured impacts. The results of this comparison are available on request.

3.A.2 Household and Location Characteristics

Table 3.A.1: List of household and location characteristics

Variable description
<i>Household demographics</i>
Household size
Age of household head
Gender of household head
Marriage status of household head
Number of children (between 0 and 4, going to primary school, going to junior high school)
Dependency ratio
Village head lives in household
<i>Education</i>
Household head's education (graduated from no school, primary school, junior high school, high school or higher)
Household member with highest qualification (graduated from no school, primary school, junior high school, high school or higher)
<i>Occupation</i>
Household head works in agricultural sector (including mining / quarrying)
Household head works in industrial sector
Household head works in service sector
<i>Housing characteristics</i>
Privately owned house
Per capita floor area
Type of floor, walls, and roof
Private toilet
Clean drinking water
Electricity source
Cooking fuel
<i>Assets</i>
Kitchen appliances (gas burner, fridge/freezer, rice cooker, mixer/blender)
Electronic devices (air conditioning, fan, radio, TV, DVD/VCD player, laptop/PC, dish antenna, cell phone)

Table 3.A.1: List of household and location characteristics

Variable description
Livestock (poultry, pig, goat, cow/buffalo, horse)
Means of transport (bike, motorbike, car)
Productive machinery (sewing machine, electric pump)
Jewelry/gold
Household ever received credit
<i>Village characteristics</i>
Number of households
Schools (primary school, junior high school)
Medical facilities (doctor, midwife, neighborhood medical center, medical center, clinic)
Semi/permanent market place
Credit facility
Road type
Mean agricultural land area
<i>Subdistrict characteristics</i>
Ratios of household heads working in agricultural / industrial / service sector
Ratios of household heads graduated from no school / primary school / junior high school / high school or higher
Mean per capita floor area
Mean agricultural land area
Ratio of households with clean drinking water

### 3.A.3 Additional Estimation Results

Table 3.A.2: Estimation results of equation 3.3.1 using only the PMT group and no imputations

Dependent variable (household level)	Satisfied	Agree with list	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Enough benefits
Targeting accuracy	1.226** (0.478)	0.834** (0.330)	-0.316 (0.374)	-1.276*** (0.310)	1.076* (0.589)	0.345 (0.539)	0.822** (0.4060)
Benefit ratio	0.474 (0.321)	0.234 (0.157)	-0.0317 (0.203)	-0.128 (0.196)	-0.100 (0.347)	0.045 (0.2900)	0.376* (0.219)
Adj. $R^2$	0.080	0.078	0.016	0.066	0.042	-0.006	0.059
Observations	383	453	456	458	342	357	460

Dependent variable (village level)	Households satisfied	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Complaints to village head (log)	Complaints in complaint box (log)	Enough benefits
Targeting accuracy	0.919 (0.567)	0.0679 (0.312)	-0.0431 (0.191)	0.502 (0.376)	0.936 (0.605)	-3.908*** (1.005)	-1.471** (0.634)	-0.159 (0.423)
Benefit ratio	1.121** (0.451)	-0.906*** (0.205)	0.319** (0.157)	0.182 (0.251)	0.211 (0.418)	-1.604** (0.724)	-1.176* (0.627)	1.147*** (0.310)
Adj. $R^2$	0.029	0.117	0.073	0.007	-0.014	0.128	0.435	0.133
Observations	206	209	209	209	208	209	209	204

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
Estimations are done by OLS. For outcomes at the household level, standard errors (in parentheses) are clustered at the village level. For outcomes at the village level, standard errors are heteroscedasticity-consistent. Households are from the PMT group. Targeting accuracy and local poverty use rank-score welfare as benchmark. All specifications control for local poverty, regional dummies, and log village size. In addition, the specifications with household-level outcomes control for dummies indicating whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of those two. The coefficients of these controls are omitted.

Table 3.A.3: Estimation results of equation 3.3.3 using survey welfare model – household level outcomes

	Satisfied	Agree with list	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Enough benefits
Targeting accuracy (rank-score welfare)	0.754 (0.504)	0.680** (0.290)	-0.650* (0.373)	-0.668** (0.327)	1.589*** (0.543)	-0.00168 (0.495)	1.105*** (0.387)
Targeting accuracy (predicted cons.)	-0.563 (0.706)	0.153 (0.392)	-0.00948 (0.413)	-0.374 (0.372)	-0.623 (0.669)	-1.059 (0.643)	-0.744 (0.517)
Local poverty rate (rank-score welfare)	0.200 (0.346)	0.294 (0.189)	0.159 (0.250)	-0.565*** (0.211)	0.211 (0.396)	0.0683 (0.402)	-0.340 (0.276)
Local poverty rate (predicted cons.)	-0.602** (0.285)	-0.26 (0.184)	0.0158 (0.201)	0.556*** (0.176)	-0.441 (0.341)	-0.626** (0.271)	0.00364 (0.236)
Benefit ratio	0.380 (0.346)	0.314** (0.155)	-0.0841 (0.201)	-0.216 (0.201)	-0.106 (0.379)	-0.098 (0.291)	0.322 (0.213)
Adj. $R^2$	0.087	0.076	0.002	0.086	0.040	0.024	0.061
Observations	383	453	456	458	342	357	460

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are clustered at the village level. Households are from the PMT group only. All specifications control for regional dummies, log village size, and dummies indicating whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of those two. The coefficients of these controls are omitted.

Table 3.A.4: Estimation results of equation 3.3.3 using survey welfare model – village level outcomes

	Households satisfied	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Complaints to village head (log)	Complaints in complaint box (log)	Enough benefits
Targeting accuracy (rank-score welfare)	0.835 (0.660)	-0.00471 (0.374)	-0.0454 (0.183)	1.047*** (0.377)	0.636 (0.705)	-0.919 (1.248)	-3.380*** (0.781)	0.0991 (0.482)
Targeting accuracy (predicted cons.)	1.136 (1.010)	-0.0444 (0.396)	0.163 (0.212)	0.233 (0.509)	-0.304 (0.878)	0.770 (1.432)	1.515 (1.058)	0.0195 (0.530)
Local poverty rate (rank-score welfare)	0.531 (0.448)	0.414* (0.230)	-0.270* (0.149)	0.746*** (0.258)	0.754* (0.450)	-0.416 (0.823)	-0.928 (0.608)	-0.736** (0.329)
Local poverty rate (predicted cons.)	-0.186 (0.409)	0.104 (0.211)	0.0963 (0.144)	-0.419* (0.231)	-0.242 (0.412)	0.646 (0.669)	0.959* (0.503)	-0.00243 (0.315)
Benefit ratio	1.277*** (0.484)	-0.975*** (0.232)	0.355** (0.168)	0.306 (0.267)	0.0727 (0.444)	-1.030 (0.814)	-1.185** (0.598)	1.256*** (0.349)
Adj. $R^2$	0.028	0.118	0.072	0.017	-0.022	0.126	0.439	0.132
Observations	206	209	209	209	208	209	209	204

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
Estimations are done by OLS. Standard errors (in parentheses) are heteroscedasticity-consistent. Households are from the PMT group only. All specifications control for regional dummies and log village size. The coefficients of these controls are omitted.

Table 3.A.5: Estimation results of equation 3.3.3 using total targeting accuracy – household level outcomes

	Satisfied	Agree with list	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Enough benefits
Targeting accuracy (rank-score welfare)	1.309** (0.560)	0.554 (0.388)	-0.427 (0.418)	-1.175*** (0.340)	1.349** (0.614)	0.519 (0.595)	0.626 (0.466)
Targeting accuracy (predicted cons.)	0.789 (1.013)	0.833* (0.469)	-0.098 (0.648)	-0.439 (0.516)	-0.515 (1.138)	-1.646 (1.015)	-0.982 (0.725)
Local poverty rate (rank-score welfare)	-0.243 (0.255)	0.177 (0.214)	0.575** (0.229)	-0.181 (0.232)	-0.417 (0.385)	-0.183 (0.347)	-0.670*** (0.240)
Local poverty rate (predicted cons.)	-0.496** (0.226)	-0.0463 (0.179)	-0.0438 (0.176)	0.241 (0.179)	-0.243 (0.274)	-0.762*** (0.230)	-0.0601 (0.210)
Benefit ratio (predicted cons.)	0.921** (0.369)	0.419** (0.205)	-0.171 (0.253)	-0.388* (0.226)	0.154 (0.462)	0.0366 (0.392)	0.335 (0.252)
Adj. $R^2$	0.088	0.067	0.020	0.053	0.046	0.054	0.059
Observations	383	453	456	458	342	357	460

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
Estimations are done by OLS. Standard errors (in parentheses) are clustered at the village level. Households are from the PMT group only. All specifications control for regional dummies, log village size, and dummies indicating whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of those two. The coefficients of these controls are omitted.



Table 3.A.6: Estimation results of equation 3.3.3 using total targeting accuracy – village level outcomes

	Households satisfied	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Complaints to village head (log)	Complaints in complaint box (log)	Enough benefits
Targeting accuracy (rank-score welfare)	0.926 (0.816)	0.556 (0.486)	-0.38 (0.351)	0.978* (0.513)	1.067 (0.909)	-2.396 (1.535)	-1.682 (1.036)	-0.986 (0.724)
Targeting accuracy (predicted cons.)	-0.111 (1.390)	0.328 (0.667)	0.549 (0.429)	-0.879 (0.751)	-2.419* (1.351)	2.976 (2.146)	0.867 (1.747)	0.125 (0.969)
Local poverty rate	0.794* (0.443)	0.539** (0.253)	-0.301** (0.134)	0.393 (0.271)	0.58 (0.448)	-0.471 (0.719)	-0.13 (0.585)	-0.920*** (0.335)
Local poverty rate (predicted cons.)	0.101 (0.375)	0.347* (0.186)	-0.0772 (0.128)	-0.0257 (0.219)	0.184 (0.346)	0.878 (0.669)	0.293 (0.500)	-0.406 (0.272)
Benefit ratio (predicted cons.)	0.968* (0.549)	-0.878*** (0.271)	0.396* (0.219)	0.128 (0.311)	-0.244 (0.527)	-1.076 (0.922)	-1.052 (0.672)	1.177*** (0.450)
Adj. $R^2$	0.024	0.142	0.100	0.007	-0.014	0.083	0.423	0.165
Observations	206	209	209	209	208	209	209	204

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
Estimations are done by OLS. Standard errors (in parentheses) are heteroscedasticity-consistent. Households are from the PMT group only. All specifications control for regional dummies and log village size. The coefficients of these controls are omitted.

Table 3.A.7: Estimation results of equation 3.3.3 including PMT and Hybrid group – household level outcomes

	Satisfied	Agree with list	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Enough benefits
Total targeting acc.	1.264*** (0.414)	0.435 (0.309)	-0.229 (0.327)	-0.313 (0.316)	1.393*** (0.539)	1.092** (0.497)	0.407 (0.420)
(rank-score welfare)							
Total targeting acc.	-0.326 (0.369)	-0.0698 (0.234)	-0.0248 (0.294)	0.0493 (0.257)	-0.575 (0.441)	-0.596* (0.332)	-0.338 (0.357)
(predicted cons.)							
Local poverty rate	0.0802 (0.255)	0.383** (0.192)	0.377 (0.251)	-0.516** (0.226)	0.0207 (0.385)	0.193 (0.352)	-0.432 (0.285)
(rank-score welfare)							
Local poverty rate	-0.366* (0.213)	-0.199 (0.166)	-0.0279 (0.174)	0.443** (0.180)	-0.253 (0.267)	-0.555*** (0.197)	0.0279 (0.223)
(predicted cons.)							
Adj. $R^2$	0.092	0.068	0.014	0.053	0.063	0.041	0.046
Observations	383	453	456	458	342	357	460

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
Estimations are done by OLS. Standard errors (in parentheses) are clustered at the village level. Households are from the PMT or Hybrid group. All specifications control for treatment group, regional dummies, log village size, and dummies indicating whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of those two. The coefficients of these controls are omitted.

Table 3.A.8: Estimation results of equation 3.3.3 including PMT and Hybrid group – village level outcomes

	Households satisfied	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Complaints to village head (log)	Complaints in complaint box (log)	Enough benefits
Total targeting acc. (rank-score welfare)	1.727** (0.682)	0.611* (0.330)	-0.513** (0.211)	0.772** (0.391)	0.916 (0.656)	-1.060 (1.220)	-2.891*** (0.947)	-1.065** (0.469)
Total targeting acc. (predicted cons.)	-0.299 (0.620)	0.0649 (0.294)	0.102 (0.160)	0.151 (0.309)	-0.412 (0.500)	0.932 (0.957)	0.247 (0.705)	-0.0806 (0.394)
Local poverty rate (rank-score welfare)	0.970** (0.484)	0.368 (0.256)	-0.301** (0.136)	0.657** (0.256)	0.834* (0.436)	-1.195 (0.735)	-0.479 (0.564)	-0.764** (0.347)
Local poverty rate (predicted cons.)	0.0637 (0.391)	0.162 (0.212)	0.0415 (0.137)	-0.0709 (0.211)	-0.0494 (0.365)	0.704 (0.632)	-0.0463 (0.487)	-0.136 (0.312)
Adj. $R^2$	0.021	0.036	0.066	0.021	-0.020	0.067	0.452	0.067
Observations	206	209	209	209	208	209	209	204

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
Estimations are done by OLS. Standard errors (in parentheses) are heteroscedasticity-consistent. Households are from the PMT or Hybrid group. All specifications control for treatment group, regional dummies, and log village size. The coefficients of these controls are omitted.



## Chapter 4

# The Impact of Road Development on Household Welfare in Rural Papua New Guinea

*Joint work with Eric Koomen, Menno Pradhan, and Christopher Edmonds*

Road access is one of the key elements necessary for rural economic development, generally offering the sole means for connecting people to markets and public services. Better roads connect rural households to goods and labor markets, providing a greater variety and lower prices of essential inputs and consumption goods, as well as higher prices and demand for local products (Gibson and Rozelle, 2003). Increased market access may raise local productivity and wages, and enable the transformation from subsistence agriculture to growing cash crops and to non-agricultural activities, thus diversifying household income sources (Mu and van de Walle, 2011; Aggarwal, 2018). Better roads may also attract financial service providers, facilitating agricultural investments and consumption smoothing (Binswanger, Khandker and Rosenzweig, 1993), and enhance access to and quality of services like schools and hospitals (Bell and van Dillen, 2020).

Each of these factors suggests that better roads lead to higher average household consumption, which is confirmed by several studies (Knox, Daccache and Hess, 2013). The distributional effects of better roads, however, are less understood, and the relative benefits to poor households are unclear. If better-off households are more able to compensate a lack of good roads—say, because they can better smooth out their consumption during seasons where roads are not usable, or through access to four-wheel drives—the poor would experience a relatively higher productivity gain from improved roads. Conversely,

non-poor households may benefit more from better roads, since they might be able to scale up agricultural production more easily, or because the poor may be kept from road utilization due to transport costs. The effects of roads on other development outcomes beside consumption are also ambiguous. Non-agricultural employment may increase or decrease as a result of better roads, depending on the relative change in agricultural productivity. School enrollment may increase or decrease, depending on how infrastructure changes education returns and the opportunity cost of schooling (Adukia, Asher and Novosad, 2020).

In this chapter, we investigate the effect of road quality on household welfare in rural Papua New Guinea (PNG) between 1996 and 2010. PNG is a good case study ground for the effects of rural infrastructure development, due to its limited road access and high variability in road quality. In 2009, the country had a road density of 56 km per 1,000 square km, which is very low compared to other countries in the region (neighboring Indonesia had 250 km per 1,000 square km). In the same year, only 13% of roads were sealed, while the majority of roads were gravel or dirt roads.

We combine two household surveys with censuses as well as GIS datasets of the road network from the time the household surveys were administered. The road map data contains measures of surface type and condition for each road section of 100 meters. Together with the location of households, we can calculate the distance to the nearest road, and the quality and length of the shortest route leading to the nearest town. This allows us to construct a set of road variables that captures quality at a very granular level, considering not just the section proximate to the households but arguably the most important road for each household in its entirety.

We use six indicators for household welfare: consumption, poverty status, household engagement in subsistence farming, wage employment, housing quality, and school enrollment. Our results suggest that access to better roads increases consumption and reduces poverty by facilitating the transition out of subsistence farming and into more market-based agriculture. Access to better roads leads to a higher probability of living in a house made of non-traditional material, and greater likelihood that children are enrolled in school. We also find that disadvantaged and more remote households benefit comparatively more from upgraded roads.

To identify the impact of road quality on the route that leads to the nearest town, we use a correlated random effects model to account for endogeneity of road quality. Following the approach by Mundlak (1978), we correct for unobserved location-specific effects by including the average of the road quality variables obtained from the two maps. The resulting estimator is identical to the within estimator, despite the absence of panel

data. Our results generally point to beneficial effects of sealing roads. The estimates suggest that, between 1996 and 2010, upgrading one kilometer of the route leading to the nearest town from dirt to sealed road surface increased average household consumption by about 3.2%, raised the chance households lived in a house with a high-quality roof by about 1.3 percentage points, reduced the probability a household relied on subsistence farming by 0.5 percentage points, and increased the likelihood for a school-aged child to be enrolled in school by 1.4 percentage points.

We also explore effect heterogeneity. Our model is run on several population subgroups to see how the effect of road quality on consumption and poverty varies by remoteness of the household, its education level, and its demographic characteristics. We find that the effects on consumption and poverty are at least twice as high for households living more than 30 km from the nearest town, when compared to those living closer than 30 km. In addition, we apply a generalized quantile regression estimator (Powell, 2020) to investigate how the effects of road infrastructure vary across the consumption distribution. This procedure allows to examine how different consumption quantiles are affected while also accounting for covariates, making the results comparable to the results from our base specification. Our estimates weakly indicate that the effect of upgrading dirt roads is higher for the poorest households, suggesting that road works can be considered anti-poverty measures in the case of rural PNG.

Our study contributes to the literature that assesses the impact of roads on rural development outcomes at the household level. A particularly relevant paper is the one by Gibson and Rozelle (2003) on PNG. Using data from the 1996 PNG Household Survey, the authors show that reducing the travel time to the nearest road to a maximum of two hours led to an overall reduction in poverty of between 5.8% and 11.8%.<sup>1</sup> Our study complements this paper by focusing on impacts of road quality rather than access.

Only a few studies have examined heterogeneity in the impacts of roads. Dercon et al. (2009) find that access to an all-weather road in rural Ethiopia between 1994 and 2004 reduced poverty by 6.9 percentage points. They find no evidence of heterogeneity of this effect across household characteristics like size of landholdings, livestock holdings, or literacy of the household head. However, their estimates show the effect on consumption growth is larger for households with landholdings of at least a hectare and a literate household head. Dercon, Hoddinott and Woldehanna (2012) obtain complementary results, finding that remoteness from towns and poor roads are among the factors most associated with chronic poverty. Khandker, Bakht and Koolwal (2009) investigate how

---

<sup>1</sup>We were able to replicate the study by Gibson and Rozelle (2003) for the new household survey data from 2009/10, and can confirm the effect of distance to the nearest road and household consumption and poverty. The replication analysis is available upon request.

households in Bangladesh profited from road improvement projects. They predict that villages next to an improved road experience a poverty reduction of 3-6 percentage points over 4 years. The impact on household expenditure is higher for lower expenditure quintiles in this study, suggesting that road investments are pro-poor. However, using a larger dataset and controlling for other investment programs, Khandker and Koolwal (2010) find the opposite pattern. Mu and van de Walle (2011) find positive and significant average effects of rural road rehabilitation on local market development in Vietnam. The authors note a tendency for poorer localities to have higher impacts due to lower levels of initial market development. A replication study by Nguyen (2019) confirms these results.

Estimating the effects of transport infrastructure is complicated by the fact that government decisions about where to construct new roads, or whether to rehabilitate or upgrade existing ones, are likely correlated with areas' growth and other development achievements. These decisions are often made based on unobserved factors like expected traffic volume, local productivity, investment cost, and political benefits of placing roads in particular areas—all factors that may also affect household welfare directly. An example of political favoritism in road placement is provided by Burgess et al. (2015), who find that during periods of autocracy in Kenya, road construction in each district was governed strongly by whether the district had a large proportion of people from the same ethnic affiliation as the current president.

Existing research in road development impacts has used a variety of approaches to address this endogeneity problem. Instrumental variable estimation—which requires an exogenous variable that affects road development but has no direct effect on the outcome variable of interest—is one approach (e.g. Gibson and Rozelle, 2003; Gertler et al., 2016; Banerjee, Duflo and Qian, 2020). A technique often used for binary road variables is difference in differences estimation, sometimes combined with propensity score matching to allow the common trend assumption to hold conditional on covariates (Lokshin and Yemtsov, 2005; Mu and van de Walle, 2011). Regression discontinuity designs have been used to study road impacts where investment decisions were made based on whether localities exceeded a threshold at some priority measure. Examples include program evaluations in Sierra Leone (Casaburi, Glennerster and Suri, 2013) and India (Aggarwal, 2018; Adukia, Asher and Novosad, 2020; Asher and Novosad, 2020).

Where panel data is available, another approach to address potential endogeneity in road works is to use village or household fixed effects (Gibson and Olivia, 2010; Khandker, Bakht and Koolwal, 2009; Khandker and Koolwal, 2010) to assess the impact of road investments over the period covered by the panel. The fixed effects approach accounts for endogeneity caused by time-invariant characteristics of the location. The availability of



multiple time periods further allows instrumentation using lagged outcomes (Dercon et al., 2009; Khandker and Koolwal, 2011). In this chapter, we apply a correlated random effects model as a way to incorporate village fixed effects. This approach is widely applicable in contexts where repeated measurements are available for the road variables but not for household data.

The remainder of the chapter is structured as follows. In section 4.1 we discuss the country context of the study and explain the data we use. Section 4.2 outlines the estimation techniques we use. In section 4.3 we present and discuss estimation results. In section 4.4, we offer some concluding points.

## 4.1 Data and Context

With a population of roughly 8.9 million people in 2020, PNG is the largest and most populated country of the Pacific region. During the years covered by the data used in this study (1996 to 2010), PNG's real per capita growth was moderate—averaging 2.6% per year—and poverty incidence even grew by two percentage points to roughly 40% of the population (Gibson, 2012). During that time period, the share of PNG's population living in rural areas has remained constant at around 87% according to the 2011 census, of which around 90% largely relied on (semi-)subsistence agriculture (see Table 4.1.1).

Development and maintenance of PNG's road network suffered during the two decades following independence, when funding for road maintenance fell by half (Kwa, Howes and Lin, 2010). Existing roads generally fell into disrepair and there was very low investment in new roads. Government expenditure on infrastructure per capita reached its minimum in 2001, but large and sustained increases in funding only began in 2010 (Dornan, 2016). Beside low government revenues, a number of other factors have made it difficult to build and maintain roads during this period: limited road management capacity in the private sector due to unsteady provision of maintenance contracts, competition for construction equipment and skilled engineers between resource extraction enterprises and the Department of Works, and disputes with owners of land proximate to road works (Lucius, 2010). Alongside these practical obstacles to implementation, a lack of political will due to the low visibility of road investments (compared to health or education spending) as well as high levels of corruption contributed to the insufficient maintenance of the road network (Dornan, 2016). Lastly, PNG's climate and geography, with its rugged topography, seismic activity, intense weathering, and high seasonal rainfall in many regions—especially in the densely populated agricultural heartland of the Highlands region—make road construction and maintenance challenging (Stead, 1990).

Two nationally representative cross-sectional household surveys conducted in PNG—the 1996 Papua New Guinea Household Survey (PNGHS 96) and the 2009–2010 Household Income and Expenditure Survey (HIES 09/10)—provide the primary source of data on household consumption and other indicators used in this study. The 1996 household survey included a nationally representative sample of 830 rural households in 73 census units (villages). The 2009 survey collected information from 2,208 rural households in 125 villages. We also make use of variables from the PNG Census 2000 Community Profile System (CPS 2000), which contains information on the location and population of all villages and towns from the census of 2000. This allows us to locate all the villages in the two surveys as points on a map.<sup>2</sup> The HIES 09/10 includes GPS coordinates of all surveyed households, allowing us to calculate fairly precise household-specific distances to the nearest road. In addition, we make use of the Papua New Guinea Resource Information System (PNGRIS) of the PNG National Agricultural Research Institute. This spatial database contains information on elevation, climate, and other biophysical characteristics which we include as control variables in our analysis.

For data on the status of road infrastructure over time, we rely on the road information data bank and the geographical information system of the Road Asset Management System (RAMS). The RAMS project, initiated in 1998 by the PNG Department of Works (DoW), was designed to provide a road database and analytical tool to inform policy makers about road needs and economic efficiency of investments in the road network (Jusi et al., 2003). We link initial RAMS data—based on road surveys conducted between 1999 and 2001—to the PNGHS 96 (we refer to the combined data as the 2000 map). Due to continued underinvestment in the transport sector, certain dimensions of the RAMS—particularly its traffic counts (vital to estimating a road’s value)—were not updated after 2001. However, the provincial works managers of the DoW were given financial support to update data on road quality, and data collected by DoW provide the basis for our dataset of roads in 2009. The road system consisted of roughly 26,000 km of roads in 2009. We link this dataset to the HIES 09/10 and refer to it as the 2009 map.<sup>3</sup>

---

<sup>2</sup>Since the PNGHS 96 used sampling based on census units from the census of 1990, on which data is unavailable, we first had to recode the 1990 census units. For this, we relied on the census unit names listed in Gibson and Rozelle (1998) as well as the generous help of staff at the National Statistical Office of PNG. The HIES 09/10 was sampled from the census of 2000, which made locating the census units straightforward.

<sup>3</sup>The two road datasets offer slightly different spatial representations of common road segments, with positional differences ranging up to several hundred meters. To ensure that our analysis is not influenced by differences in the coverage and spatial representation of the road network across the two years, we use the representation of the 2009 map for both maps. That means that we have to include the information on surface type and road condition of 2000 in the 2009 road map. We match roads based on road section IDs, and where those are missing, on spatial proximity. It should also be mentioned that we cannot make use of the most detailed survey of the national road network to date, the Comprehensive Visual

Both the 2000 and 2009 road maps include detailed information for each road segment. Most importantly, surface type (sealed, gravel, dirt) and condition (good, fair, poor) are given for all segments.<sup>4</sup> The 2009 road map has much more extensive coverage of the road network—covering an additional 14,000 km of roads not included in the 2000 map, consisting mostly of gravel or dirt feeder roads. However, within our study period the focus of road works in PNG was on maintenance and upgrading, and reportedly no new roads were constructed during this period (World Food Programme and Logistics Cluster, 2011; also confirmed in discussions with several informants at the DoW). This leads us to conclude that the higher density of roads depicted in the 2009 map is a result of an improvement of the information contained in the map, rather than the construction of new roads.

Figures 4.1.1 maps the roads, distinguished by surface type, comprising identified stretches of the national network in 2000 and 2009. A comparison of the maps shows that most of the missing road segments in 2000 are classified as dirt roads in 2009. This is in part due to the fact that the additional roads on the 2009 map are made up almost entirely of provincial roads, which were more likely than national roads to have a dirt surface. Section 4.A.1 in the appendix contains descriptive statistics of the road network for both years, as well as transition matrices of surface type and condition for those roads included in both maps. They do not reveal an apparent trend in road development, in that the length of roads upgraded or improved is roughly offset by roads that deteriorated.

Using data from the road maps and household surveys, we construct variables indicating the length, surface types and condition of the road leading to the nearest town for the sampled households. We consider the shortest route to the nearest town from the stretch of road that is closest to the household. This route may not be a perfect representation for the roads used by surveyed households—as sometimes the nearest town may not be the most relevant—but we believe that it is a useful heuristic nonetheless. For this route, we calculate the distance for each type of surface and road condition.<sup>5</sup> For our analysis, we focus on households that are connected to a town by a road. We exclude households

---

Road Condition Survey, which was collected in 2014 by the Papua New Guinea-Australia Transport Sector Support Program (TSSP) together with the DoW. Due to the sudden heavy rise in national road investments starting in 2011, we believe that the conditions in this new survey do not adequately reflect the conditions around the time the HIES 09/10 was conducted.

<sup>4</sup>Guidelines for the interpretation of the classifications of road condition are provided in the CAPE-PNG-9-Transport-Sector-Assessment (<https://www.adb.org/sites/default/files/linked-documents/CAPE-PNG-9-Transport-Sector-Assessment.pdf>). A road segment is labeled “good” if it is passable for a two-wheel vehicle in wet weather, “fair” if it requires periodic maintenance, and “poor” if it requires reconstruction or rehabilitation.

<sup>5</sup>Details on the construction of this variable, as well as a discussion of remote sensing as an alternative to identify road quality and surface type, can be found in the appendix of Edmonds et al. (2018).

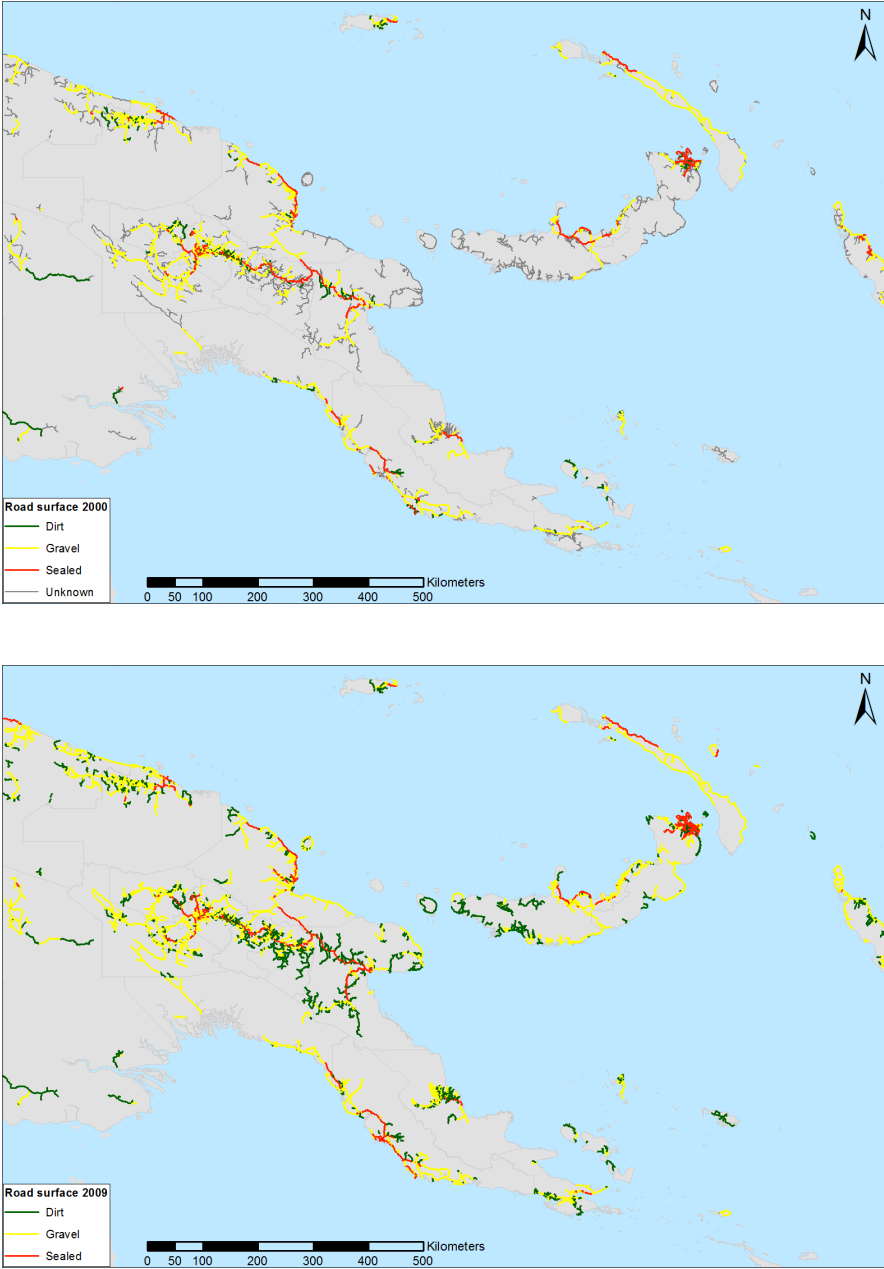


Figure 4.1.1: Roads by surface type in 2000 and 2009

that are located more than 15 km from any road as it seems unlikely that they use the nearest road regularly, making road quality unlikely to have a significant impact. Furthermore, we consider only towns that had more than 1,000 inhabitants according to the census 2011 and are within 5 km of the nearest road.<sup>6</sup> Taken together, these restrictions mean that about 20% of the rural census units in the two surveys are dropped from our analysis. The villages left out are mostly either on small islands (where roads may be of little or no importance anyway), or in the deep interior in the west of the Momase region or at the coast of Western province (both of which have a very low population density).

Due to the fact that our road map of 2000 is less detailed than the 2009 map, we have missing information on surface type and condition for some of the road segments of 2000. For the routes used in the analysis, this information was lacking for 24% of the total distance. For that reason, we drop observations where all road segments leading to the nearest town have unknown characteristics for 2000. This leaves a total of 20% of the total distance unknown for 2000. For the remaining observations, we simply assign the segments with unknown characteristics for 2000 the same characteristics as for 2009. We discuss the implications of this decision in section 4.3.

Table 4.1.1 provides summary statistics for our sample. It includes geographic and road access variables that were merged with the household survey data. Both household surveys include sections that allow calculation of per adult-equivalent household expenditure—based on a closed interval recall method in the PNGHS 96, and on consumption diaries in the HIES 09/10—as well as regional poverty lines based on the cost of locally consumed foods.<sup>7</sup> Across the two household surveys, average per adult-equivalent consumption decreased and poverty incidence increased slightly. School enrollment increased by more than 14 percentage points, and average years of schooling of adults increased by almost one year. Households in the 2009/10 HIES were slightly older (higher average age of household members), smaller (with nearly one person less, on average), and lived closer to the nearest town, compared to the households in the PNGHS 96. These differences could indicate demographic changes, but are more likely caused by differences in the sampling schemes for the two surveys. Table 4.1.1 also shows the distances by

---

<sup>6</sup>The latter criterion leads to the exclusion of three settlements that can only be reached by water. All towns with more than 1,000 inhabitants in 2011 were already towns in 2000.

<sup>7</sup>We construct per capita expenditure as well as regional poverty lines as explained in Gibson and Rozelle (1998) and Gibson (2012). Particularly, we use the revised consumption figures, poverty lines, and sampling weights for the PNGHS 96 explained in Gibson (2012) to make expenditure and poverty comparable between the two surveys. For the HIES 09/10, Gibson (2012) suggests three different consumption figures. Due to evidence of diary fatigue, we use the figure based on the shortest time horizon (7 days). The poverty lines we use take the cost of a locally consumed food basket and add the non-food spending of households whose food expenditures exactly meet this cost (Olson Lanjouw and Lanjouw, 2001).

Table 4.1.1: Summary statistics for the analysis sample

	1996			2009/10			<i>p</i> -value
	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	
<i>Outcomes (PNGHS 96, HIES 09/10)</i>							
Log real per adult-equivalent expenditure	0.309	0.560	680	0.157	0.946	1,473	0.033
Poverty	0.347	0.376	680	0.429	0.583	1,473	0.045
Engaged in subsistence farming	0.889	0.253	680	0.914	0.318	1,643	0.399
Someone in the household has wage job	0.243	0.346	680	0.137	0.390	1,632	0.003
Home has a good roof	0.244	0.346	680	0.285	0.512	1,643	0.205
Ratio of children aged 7-17 going to school	0.445	0.325	484	0.589	0.491	1,054	0.000
<i>Location-specific control variables (PNGRIS)</i>							
Altitude (m)	1003	595	680	1185	951	1,643	0.013
Dummy, slope > 10 degrees	0.634	0.388	680	0.664	0.536	1,643	0.718
Dummy, land inundation occurs	0.236	0.342	680	0.240	0.485	1,643	0.955
Dummy, rainfall deficit is rare	0.290	0.366	680	0.322	0.530	1,643	0.695
Annual rainfall (m)	2.640	0.513	680	2.650	0.813	1,643	0.929
Log population nearest town (Census)	8.615	1.007	680	8.805	1.407	1,643	0.374
<i>Household controls (PNGHS 96, HIES 09/10)</i>							
Household size	5.839	2.438	680	4.906	2.785	1,643	0.000
Ratio of household members under 15	0.406	0.208	680	0.352	0.263	1,643	0.000
Ratio of household members above 50	0.104	0.172	680	0.127	0.268	1,643	0.117
Age of household head	40.75	10.32	680	42.47	15.61	1,643	0.032
Household head is female	0.094	0.235	680	0.148	0.402	1,643	0.006
Household head is literate	0.525	0.403	680	0.473	0.567	1,642	0.192
Average years of schooling	3.392	2.389	680	4.373	3.993	1,633	0.000
Ratio of children in primary school age	0.602	0.302	484	0.583	0.433	1,054	0.469
<i>Impact variables (obtained from road maps)</i>							
Distance to nearest road (km)	0.609	1.234	680	0.907	2.141	1,643	0.262
Distance on road to nearest town (km)	49.99	39.03	680	36.57	39.07	1,643	0.044
Road to town: km of sealed road	16.10	21.15	680	14.68	28.48	1,643	0.751
Road to town: km of gravel road	30.47	27.67	680	17.69	24.67	1,643	0.007
Road to town: km of dirt road	3.428	6.206	680	4.199	10.92	1,643	0.589

Summary statistics of log real per adult-equivalent expenditure and poverty are obtained using person sampling weights. All other ones are obtained using household sampling weights. Standard deviations and *p*-values account for clustering at the village level. The *p*-values show the probability of equal means.

surface type for the routes leading from the sampled households to the nearest town, as measured by us using the road maps and household locations.

We use six different outcome variables in our regressions of road quality. The first one is the same one as used in Gibson and Rozelle (2003): the logarithm of real yearly consumption per adult-equivalent.<sup>8</sup> We divide yearly consumption by the respective regional rural poverty lines to calculate real consumption. A similar outcome variable is poverty status. We include it as our second outcome to examine specifically how the probability of being poor is affected by road infrastructure.

In addition to consumption, we are also interested in the effects of infrastructure on rural employment and structural transformation. One common change observed among rural households as a result of improved access to markets is reduced dependence on subsistence farming. Therefore, our third outcome is whether members of the household reported engaging in subsistence farming in the days prior to the survey date.<sup>9</sup> Another indicator of structural change is having a wage job. Accordingly, we use a dummy for whether at least one member of the household is employed in a non-agricultural sector as fourth outcome variable. We expect that better roads improve off-farm earning opportunities and therefore reduce the necessity for subsistence agriculture and increase the likelihood of formal employment.

Another indicator of material well-being is housing quality.<sup>10</sup> Traditional dwellings, with walls made of bush material and with grass or leaf roofs, are cheap to come by but have a relatively low service life. So, investments in non-traditional housing materials may signify not only an immediate improvement of living conditions but also an intention to stay in a given location more permanently. Due to the lack of credible and intertemporally comparable housing value estimates, we chose to use a dummy indicating whether the house has a good roof, i.e., a roof made of metal, tiles, or cement. If improved road access leads to better opportunities and living conditions and reduces transportation costs, we would expect more good roofs due to better access to roofing materials as well as higher demand for them.

Lastly, we examine the school enrollment ratio of children at school age. Our hypoth-

---

<sup>8</sup>Like Gibson and Rozelle (2003), we assign children aged between 0 and 6 years a weight of 0.5, while children older than 6 years as well as adults are assigned a weight of 1.

<sup>9</sup>The definitions vary slightly between survey rounds. For the PNGHS 96, the variable indicates that in the two weeks prior to the survey, at least one household member engaged in the production of sago, bananas, corn, sweet potato, cassava, taro, or other fresh fruits or vegetables without selling them. For the HIES 09/10, the variable indicates that in the week prior to the survey, at least one household member engaged in agricultural production for own consumption. The means of both variables are very close, as shown in Table 4.1.1.

<sup>10</sup>Measures of housing quality were left out in the construction of the consumption figures (Gibson, 2012), so are not part of the first two outcome variables.

esis is that school enrollment increases with better road infrastructure due to easier access to schools for both children and teachers.

## 4.2 Estimation

To estimate the causal relationship between the quality of the road infrastructure and our outcomes of interest, we propose a linear model of the form

$$y_{ijt} = \beta R_{jt} + \gamma D_j + \delta X_{ijt} + \mu_j + \pi_{pt} + \varepsilon_{ijt}, \quad (4.2.1)$$

where  $y_{ijt}$  is an outcome for household  $i$  in village  $j$  at time  $t$ ,  $R_{jt}$  and  $D_j$  are vectors of variables related to road infrastructure,  $X_{ijt}$  is a vector of exogenous control variables (at the household- and the village level),  $\mu_j$  denotes unobserved, time-invariant heterogeneity at the village level,  $\pi_{pt}$  is a province-time-fixed effect, and  $\varepsilon_{ijt}$  is an independent disturbance term.

We include two types of road infrastructure variables.  $D_j$  includes the Euclidean distance to the nearest road and the distance on that road leading to the closest town with a 2010 population above 1,000 people. Since we assume no new roads were added between 1996 and 2010, these distance variables are time-invariant.  $R_{jt}$  captures the combined length of road segments of a particular type on the route between surveyed households and the nearest town. Since these distances add up to the total distance traveled on that road to the nearest town, which is already included in  $D_j$ , we leave out the lowest quality category of road type. This means that the coefficients in  $\beta$  give the welfare impact of changing a kilometer of road from the lowest quality type to the other types. Road segments are upgraded, left to deteriorate, or remain the same type over time, so  $R_{jt}$  is time-varying.

Since there is no overlap in sampled villages between the two surveys, we cannot difference out the term  $\mu_j$ , and treating it as a random effect uncorrelated with all independent variables might lead to a biased estimate of  $\beta$ . Instead, we use the correlated random effects approach introduced by Mundlak (1978). The idea is to substitute  $\mu_j$  with the mean of  $R_{jt}$  over  $t$ ,  $\bar{R}_{j*}$ , and an independent village random effect,  $\omega_j$ :

$$\mu_j = \alpha \bar{R}_{j*} + \omega_j. \quad (4.2.2)$$

$R_{jt}$  is uncorrelated with  $\omega_j + \varepsilon_{ijt}$ . This combined disturbance term is also independent between nearby localities, as any spatial correlation is captured by  $\pi_{pt}$  and  $\bar{R}_{j*}$ . Since distance variables  $D_j$  are time-invariant, they cannot be included in the model of  $\mu_j$  and



are assumed to be conditionally exogenous. The model is estimated using OLS and takes into account the survey design, i.e., sampling weights<sup>11</sup>, stratum divisions, and standard error clustering by village. To reduce collinearity between the endogenous variables and their respective intertemporal averages, we estimate equation (4.2.1) using centered terms  $R_{jt} - \bar{R}_{j*}$  instead of  $R_{jt}$  to reflect variables in terms of their deviations from averages over time.

We include a number of location-specific control variables. Among these are the geoclimatic variables used in Gibson and Rozelle (2003)—namely, altitude (in meters), a dummy for whether the slope is above 10 degrees, a dummy indicating that the land is subject to flooding, a dummy indicating that rainfall deficits are rare, and annual rainfall (in meters). To control for the economic importance of the nearest town, we also include the logarithm of its population as measured in the census closest to the survey year.<sup>12</sup> Population numbers could be endogenous, say, because more productive areas could enhance household welfare as well as faster population growth. To account for this potential source of endogeneity, we also include the average of log population in equation (4.2.2).

We also include some household-level controls. This demands caution, however, since changes in road access could alter household characteristics. An example for such an endogenous household variable is the sector of work, which could change as a result of new opportunities created by improved road access, and thus cannot be included. We select a parsimonious set of variables that describes the composition and education level of adults in the household (see Table 4.1.1). We report results with and without these household controls for equation (4.2.1). Province-time fixed effects are included to control for differences in outcomes due to unobserved factors such as local economic conditions, and the ability and political will to build, maintain, and upgrade roads. They constitute particularly effective controls since the roads to the nearest town, while not having geographical point locations, are for almost all observations located within the same province or at least have the largest part in the province as the corresponding villages.

With regard to road types, we explore specifications with different levels of detail. A simple way to capture road quality is to consider only the surface type, i.e., whether a road is sealed, graveled, or a dirt track. A more detailed categorization of road segments

---

<sup>11</sup>The sampling weights used in the regressions of consumption and poverty status are person-specific, those in the regressions of schooling are children-specific, and those in the other regression are household-specific.

<sup>12</sup>We take the population size for 1996 from the 2000 census, and the population size for 2009/10 from the 2011 census.

includes surface type together with road condition, i.e., whether a road is in good, fair, or poor shape. In this chapter, the focus is on the results of the analysis using only surface type. The more detailed model is consistent with the simpler model but shows no clear effects of road condition, likely due to a lack of statistical power as well as differences in the way road condition was assessed across time and between provinces. The results using road condition can be found in section 4.A.2 in the appendix.

Our model specification rests on the assumption that all unobserved factors below the province level which both contribute to the respective outcomes and are correlated with road infrastructure are location-specific and fixed over the period between the two surveys. This may look like a daring presumption given the 13-year period between surveys. Although overall rural economic output and poverty have stagnated over the study period, some areas may have gained or lost in population or economic importance in those years, potentially affecting infrastructure as well as household welfare. We attempt to capture changes in local economic conditions by including the province-time dummies and town population numbers. With these precautions, we believe that time-invariant cluster-level heterogeneity is a reasonable assumption.

A caveat is that we cannot model selective migration that occurs in response to changes in road quality. This is a potential source of bias for most studies on the benefits of infrastructure; even when panel data is available, the whereabouts of individual migrating household members are seldom recorded. The actual effects of rural road quality on migration are unclear, with some evidence pointing to reduced out-migration (Castaing Gachassin, 2013), while other recent studies find no significant effects (Aggarwal, 2018; Asher and Novosad, 2020). In section 4.A.3 in the appendix, we look at correlations between our road variables and indicators of migration in PNG, which all turn out small and statistically insignificant. Given this, we believe that selective migration does not bias our estimates to a degree that would invalidate them.

In addition to the average welfare effects of rural roads, we are also interested in whether rural roads affect all households in the same manner. One open question is whether high education levels are complementary to road infrastructure, or whether it is mostly low-skilled labor that becomes more productive through better roads. Other sources of effect heterogeneity might be the gender and age of household members. For example, additional opportunities created by roads may help empower women and thereby have a larger effect on their consumption. Poor quality or lacking road infrastructure may trap older people and diminish their prospects more than those of young people due to physical constraints on walking long distances or transiting rough roads. The marginal effects of road quality may also differ between households who live close to a town and

those who do not. To investigate whether this type of impact heterogeneity exists, we follow Dercon et al. (2009) and divide our sample in two subsamples to estimate the model separately by subsample. We define the subsamples on the basis of: (i) whether the road distance to the nearest town is more than 30 km, (ii) whether the average years of education for household members at least 18 years old is larger than 4 years, (iii) whether the household head is female, and (iv) whether the ratio of household members above 50 exceeds 30%.

Last of all, we explore effect heterogeneity across the distribution of real consumption, which the impact estimates on poverty status do not capture. The effect of roads on poverty is driven by relatively few households around the poverty line. However, we are also interested in how relatively poor households are affected by infrastructure compared to relatively well-off households. To this end, we employ quantile regressions.

The principal idea of quantile regression is that unlike in the least squares regression framework, it is not the conditional expectation but a conditional quantile of the outcome that is a linear function of the covariates. The  $\tau^{\text{th}}$  quantile of consumption  $y_{ijt}$  conditional on road quality  $R_{jt}$  is expressed in the linear quantile function  $q(R_{jt}, \tau) = \beta(\tau) R_{jt}$ . Each possible outcome can be related to this function:

$$y_{ijt} = q(R_{jt}, U_{ijt}^*), \quad U_{ijt}^* \sim \mathcal{U}(0, 1), \quad (4.2.3)$$

where  $U_{ijt}^*$  is a non-separable disturbance term normalized to a standard uniform distribution.  $U_{ijt}^*$  determines the rank of the outcome within the conditional distribution, and is what causes heterogeneity in outcomes conditional on  $R_{jt}$ .

Khandker, Bakht and Koolwal (2009) and Khandker and Koolwal (2010) use quantile regressions in their studies of road infrastructure. Similarly to our estimation above, they include correlated random effect terms as covariates in their quantile regressions to account for unobserved heterogeneity. The issue with this approach is that including control variables in a quantile regression model alters its interpretation. For instance, the  $\tau^{\text{th}}$  consumption quantile for households with good roads is not the same as the  $\tau^{\text{th}}$  quantile for households with good roads *and* low levels of education. Using a similar reasoning, including the correlated random effect terms—and thereby, implicitly, an approximation to the location-specific fixed effect—yields an interpretation different from the one associated with model (4.2.3). On the other hand, not conditioning on controls and the correlated random effect terms likely creates biased estimates. It requires the assumption that  $U_{ijt}^*$  is independent of  $R_{jt}$ , which is quite strong, and likely holds only conditionally on covariates.

To circumvent this problem, we apply a generalized quantile regression (GQR) model,

as introduced by Powell (2020). Here, outcomes are modeled by the same quantile function as in (4.2.3), but the disturbance term  $U_{ijt}^*$  is itself dependent on covariates  $C_{ijt}$ . GQR jointly estimates  $\beta(\tau)$  and the conditional quantile  $P(y_{ijt} \leq R_{jt}\beta(\tau) | C_{ijt})$ . The covariates are thus used to predict the position in the conditional outcome distribution. This way, including control variables in the model does not alter the causal interpretation of the quantile impact estimates of road quality.

For our model, we focus on the distances by surface type  $R_{jt}$ , while the time-invariant distances to the nearest road and to the nearest town,  $D_j$ , are included in the set of covariates  $C_{ijt}$ . The main reason is that the more variables are conditioned on, the lower is the variance of the conditional outcome distributions, and the lower is the difference between quantile effects. So, if we considered the outcome distribution conditional on both  $R_{jt}$  and  $D_j$ , we would not expect to see much effect heterogeneity. As further control variables we include everything we controlled for in our most detailed specification above, which includes location- and household-specific control variables, province-time dummies, and the Mundlak terms  $\bar{R}_{j*}$ . For estimation, we use the Stata routine `genqreg` for all quantiles from 1 to 99. To obtain 95% confidence intervals we apply a cluster bootstrap at the village level.

### 4.3 Results

Our main estimation results, based on equation (4.2.1), are summarized in Tables 4.3.1 to 4.3.3. For each outcome variable, we estimate three specifications. Specification 1 is a linear regression of the model specified in equation (4.2.1), without household level controls and without Mundlak terms. Specification 2 includes household-level control variables. Specification 3 in addition contains the Mundlak terms. The first two specifications allow to examine the effect that time-varying control variables have on the main coefficients. As these lead to considerable improvements in fit while leaving the coefficients of road-related variables nearly unchanged, we are not worried about them being endogenous to the model. Furthermore, the Mundlak terms in specification 3 are jointly statistically significant for four of the six outcomes. Accordingly, the third specification is our preferred one. For the analyses by subgroups and the quantile regressions we use this one as well.

To account for the fact that we test multiple hypotheses, we compute sharpened False Discovery Rate (FDR)  $q$ -values using the two-stage procedure described in Anderson (2008), in addition to  $p$ -values. The  $q$ -values (in square brackets) indicate the smallest level at which the coefficient in question would be significant, given the  $p$ -values of the simultaneously tested hypotheses. We apply this correction using all road quality variables

and all outcomes together, but separately for each specification.

Table 4.3.1 shows the impact of road condition variables on log real per adult-equivalent consumption and poverty status. The results indicate that improved road conditions have a positive impact on consumption. An upgrade of one kilometer of dirt road to sealed road on the route to the nearest town leads to a 3.2% increase in household consumption. Upgrading one kilometer of gravel road to sealed road increases consumption by 2.2%. The difference between dirt and gravel roads is positive but not significant. Similarly, transforming a kilometer of dirt road to sealed road reduces the probability to be poor by 1.3 percentage points, an upgrade to gravel road reduces it by about 1.1 percentage points. The effects of sealing dirt roads on consumption and poverty hold up to the scrutiny of FDR  $q$ -values, while the effects of graveling dirt roads have  $q$ -values slightly above 10%.

Table 4.3.2 presents estimates of the impact of road type on the likelihood of engagement in subsistence farming and in wage employment. The point estimates of the effects on these two outcomes indicate that better quality roads facilitate the structural transformation from subsistence farming to economic activities that are more integrated into local markets. In particular, a one kilometer change of road surface from gravel to sealed reduces the probability that a member of the household engages in subsistence farming by around 0.6 percentage points. For wage employment, the picture is less clear. It appears that turning gravel into sealed roads actually reduces the probability that a household member has a wage job. But the Mundlak terms in specification 3 are jointly insignificant with a  $p$ -value of 0.93, so including them might only drive up coefficient variability without reducing bias. Indeed, when leaving the Mundlak terms out, the effects break down to zero. In summary, better roads drive people out of subsistence farming and presumably toward cultivation of cash crops, as opposed to non-agricultural employment.

Table 4.3.3 shows the effects on the likelihood of having a good roof and of school enrollment of children between 7 and 17. We find clear signs that improvements in roads lead to investment in housing. A one kilometer increase in gravel versus dirt roads raises the probability of having a good roof by around 1.3 percentage points. The same holds for sealed versus dirt roads, even though the latter result is marginally insignificant. These findings are unsurprising given the high transportation costs of tiles or corrugated sheet metal. Finally, upgrading one kilometer of dirt road on the route to the nearest town to sealed or gravel appears to increase the probability of a school-aged child to be enrolled in school by 1.4 and 1.6 percentage points, respectively. However, the FDR  $q$ -values for this last result are slightly above the 10% significance threshold, making it a somewhat speculative finding.

The magnitude of some of the effects per kilometer of changed surface type appears

Table 4.3.1: Impact of road type and distances on consumption and poverty status

	Log(real p.a.e. consumption)			Poverty Status		
	(1)	(2)	(3)	(1)	(2)	(3)
Dirt to Sealed (km)	0.0004 (0.921)	0.0000 (0.994)	0.0317*** (0.008)	-0.0018 (0.386)	-0.0016 (0.446)	-0.0131** (0.020)
	[1.000]	[1.000]	[0.050]	[1.000]	[1.000]	[0.063]
Dirt to Gravel (km)	-0.0022 (0.579)	-0.0032 (0.412)	0.0217* (0.051)	0.0007 (0.708)	0.0012 (0.531)	-0.0107* (0.060)
	[1.000]	[1.000]	[0.105]	[1.000]	[1.000]	[0.109]
Gravel to Sealed (km)	0.0026 (0.176)	0.0033* (0.084)	0.0100 (0.155)	-0.0025** (0.022)	-0.0028** (0.011)	-0.0024 (0.501)
	[0.886]	[0.459]	[0.157]	[0.131]	[0.107]	[0.287]
Total distance to nearest town (km)	-0.0027 (0.456)	-0.0027 (0.464)	0.0022 (0.570)	0.0024 (0.169)	0.0024 (0.175)	0.0002 (0.913)
	[0.0069]	[0.0177]	[0.0075]	[0.0071]	[0.0128]	[0.0082]
Distance to nearest road (km)	(0.690)	(0.236)	(0.601)	(0.495)	(0.157)	(0.395)
Household controls	No	Yes	Yes	No	Yes	Yes
Mundlak terms included	No	No	Yes	No	No	Yes
$R^2$	0.188	0.315	0.321	0.131	0.213	0.215
Villages	155	155	155	155	155	155
Households	2153	2153	2153	2153	2153	2153
$p$ -value Mundlak			0.016			0.005

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

$p$ -values in parentheses, using standard errors clustered at the village level. FDR  $q$ -values in square brackets. All regressions are weighted using person sampling weights from both surveys. Dirt road is the excluded category for the distance variables. The coefficient for “Gravel to Sealed” is not part of the model and comes about by subtracting “Dirt to Gravel” from “Dirt to Sealed”. Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications also include location-specific control variables (see table 4.1.1) as well as province-time-fixed effects. Columns labeled (2) in addition include the household-level control variables (see table 4.1.1), and columns labeled (3) include Mundlak terms and household-level control variables. The last row displays the  $p$ -value of a Wald test of the joint significance of the Mundlak terms.

Table 4.3.2: Impact of road type and distances on subsistence farming and wage employment

	Subsistence farming household			Someone in household has wage job		
	(1)	(2)	(3)	(1)	(2)	(3)
Dirt to Sealed (km)	0.0010 (0.248) [0.886]	0.0011 (0.229) [0.901]	-0.0047 (0.115) [0.151]	0.0012 (0.402) [1.000]	0.0000 (0.992) [1.000]	-0.0088 (0.161) [0.157]
Dirt to Gravel (km)	0.0002 (0.849) [1.000]	0.0005 (0.626) [1.000]	0.0009 (0.716) [0.289]	0.0018 (0.168) [0.886]	0.0008 (0.445) [1.000]	0.0049 (0.286) [0.226]
Gravel to Sealed (km)	0.0008 (0.188) [0.886]	0.0006 (0.265) [0.901]	-0.0057*** (0.006) [0.050]	-0.0006 (0.517) [1.000]	-0.0008 (0.361) [1.000]	-0.0137*** (0.004) [0.050]
Total distance to nearest town (km)	-0.0004 (0.622)	-0.0005 (0.505)	-0.0006 (0.550)	-0.0027** (0.016)	-0.0016* (0.083)	-0.0014 (0.352)
Distance to nearest road (km)	0.0056 (0.210)	0.0053 (0.213)	0.0055 (0.265)	-0.0386*** (0.000)	-0.0324*** (0.000)	-0.0353*** (0.000)
Household controls	No	Yes	Yes	No	Yes	Yes
Mundlak terms included	No	No	Yes	No	No	Yes
R <sup>2</sup>	0.148	0.192	0.196	0.128	0.220	0.226
Villages	155	155	155	155	155	155
Households	2323	2322	2322	2312	2311	2311
p-value Mundlak			0.029			0.932

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
p-values in parentheses, using standard errors clustered at the village level. FDR q-values in square brackets. All regressions are weighted using household sampling weights from both surveys. Dirt road is the excluded category for the distance variables. The coefficient for “Gravel to Sealed” is not part of the model and comes about by subtracting “Dirt to Gravel” from “Dirt to Sealed”. Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications also include location-specific control variables (see Table 4.1.1) as well as province-time-fixed effects. Columns labeled (2) in addition include the household-level control variables (see Table 4.1.1), and columns labeled (3) include Mundlak terms and household-level control variables. The last row displays the p-value of a Wald test of the joint significance of the Mundlak terms.

Table 4.3.3: Impact of road type and distances on having a good roof and school enrollment

	Home has a good roof			Ratio of children going to school		
	(1)	(2)	(3)	(1)	(2)	(3)
Dirt to Sealed (km)	0.0042** (0.012)	0.0038** (0.023)	0.0133 (0.143)	0.0022 (0.326)	0.0012 (0.554)	0.0143* (0.081)
	[0.113]	[0.140]	[0.157]	[0.960]	[1.000]	[0.133]
Dirt to Gravel (km)	0.0052*** (0.002)	0.0048*** (0.004)	0.0130** (0.016)	0.0009 (0.710)	-0.0004 (0.869)	0.0155** (0.046)
	[0.038]	[0.084]	[0.063]	[1.000]	[1.000]	[0.105]
Gravel to Sealed (km)	-0.0010 (0.313)	-0.0010 (0.271)	0.0004 (0.954)	0.0013 (0.250)	0.0016 (0.157)	-0.0012 (0.771)
	[0.960]	[0.901]	[0.361]	[0.886]	[0.787]	[0.294]
Total distance to nearest town (km)	-0.0049*** (0.002)	-0.0044*** (0.005)	-0.0029 (0.149)	-0.0014 (0.528)	-0.0004 (0.862)	0.0020 (0.239)
Distance to nearest road (km)	-0.0299*** (0.000)	-0.0261*** (0.001)	-0.0298*** (0.000)	-0.0222** (0.012)	-0.0129* (0.094)	-0.0205** (0.017)
Household controls	No	Yes	Yes	No	Yes	Yes
Mundlak terms included	No	No	Yes	No	No	Yes
$R^2$	0.453	0.468	0.470	0.182	0.244	0.250
Villages	155	155	155	154	154	154
Households	2322	2321	2321	1538	1537	1537
$p$ -value Mundlak			0.194			0.079

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

$p$ -values in parentheses, using standard errors clustered at the village level. FDR  $q$ -values in square brackets. The regressions of having a good roof are weighted using household sampling weights from both surveys, the ones of schooling use child sampling weights. Dirt road is the excluded category for the distance variables. The coefficient for "Gravel to Sealed" is not part of the model and comes about by subtracting "Dirt to Gravel" from "Dirt to Sealed". Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications also include location-specific control variables (see Table 4.1.1) as well as province-time-fixed effects. Columns labeled (2) in addition include the household-level control variables (see Table 4.1.1), and columns labeled (3) include Mundlak terms and household-level control variables. The last row displays the  $p$ -value of a Wald test of the joint significance of the Mundlak terms.



quite large, given a median of 29 km to the nearest town for our sample. This can in part be attributed to the long observation period of 14 years, which may have compounded the differences in outcomes between households with better and worse roads. More importantly, the estimates are identified over a road system with relatively few changes over the years (see the transition matrix in Table 4.A.2 in the appendix). The marginal effects of those changes are likely somewhat diminishing: road planners can be expected to maintain and upgrade road segments with an eye to necessity or expected gains. Our estimates are thus not the effect of random changes in road quality, but of the kind of changes that actually occurred during the study period. The high magnitude of our findings seems to confirm the underfunded state of PNG's road policy at the time.

Another reason for a possible bias of our estimates away from zero lies in the way we handle road segments with missing road surface in the 2000 map. Since we assign the same surface type as the corresponding segments in the 2009 map to the missing segments, it is possible that actual changes along the route are biased downwards, leading to an overestimation of marginal effects. We repeat all the estimations but gradually exclude households with more than 50%, 20%, 5%, and 0% of segments missing on their shortest route to the nearest town, thereby respectively reducing the estimation sample to about 85%, 63%, 42%, and 20% of its original size. For all the outcomes and all but the last sample reduction, the results remain qualitatively the same and remain statistically significant, with most estimates even further away from zero than the original. Regression tables are omitted here but are available on request. The exercise indicates that our treatment of missing segments does not have a major effect on the results.

Next, we discuss the estimates disaggregated by subgroups of households to study heterogeneous effects of road quality. We only consider the models of consumption and poverty, since these are the key outcome variables of this chapter. The results are reported in Tables 4.3.4 and 4.3.5, with the subsample defined in the top row. In the third row from the bottom of the tables, we report estimated correlation coefficients between the outcome variables and subset identifiers. In the last two rows, we provide  $p$ -values of Wald tests of the equality of the surface material coefficients between the two subsamples, and of the same for the coefficients of all road variables, respectively.

We find clear differences when effects are estimated for households distinguished by distance to the nearest town. Households living farther than 30 km to the nearest town benefit more from dirt road upgrades than those living closer, both in terms of consumption as well as the likelihood of being poor. It appears that for shorter trips road quality plays a minor role, while for longer distances the travel cost differences between better and worse roads may prove pivotal for the decision to make a trip more often, with measurable

Table 4.3.4: Impact of road type and distances on log real per adult-equivalent consumption by subgroups

	Road distance to nearest town $\leq$ 30km	Road distance to nearest town $>$ 30km	Average years of schooling $\leq$ 4	Average years of schooling $>$ 4	Household head male	Household head female	Household members above 50 $\leq$ 30%	Household members above 50 $>$ 30%
Dirt to Sealed (km)	0.0107 (0.801) [0.560]	0.0365*** (0.004) [0.052]	0.0294* (0.080) [0.189]	0.0248 (0.133) [0.230]	0.0303*** (0.006) [0.052]	0.0463 (0.224) [0.265]	0.0188 (0.113) [0.223]	0.0460** (0.011) [0.070]
Dirt to Gravel (km)	-0.0476* (0.076) [0.189]	0.0378*** (0.001) [0.033]	0.0240** (0.049) [0.174]	0.0121 (0.461) [0.361]	0.0206** (0.049) [0.174]	0.0250 (0.146) [0.230]	0.0160 (0.110) [0.223]	0.0295* (0.081) [0.189]
Gravel to Sealed (km)	0.0583* (0.055) [0.180]	-0.0014 (0.867) [0.560]	0.0054 (0.664) [0.553]	0.0126 (0.137) [0.230]	0.0098 (0.127) [0.230]	0.0213 (0.548) [0.438]	0.0028 (0.761) [0.560]	0.0166* (0.080) [0.189]
Total distance to nearest town (km)	-0.0076 (0.514)	0.0066 (0.237)	0.0061 (0.163)	-0.0073 (0.214)	0.0029 (0.428)	0.0065 (0.430)	0.0004 (0.928)	0.0064 (0.225)
Distance to nearest road (km)	0.0598* (0.089)	-0.0074 (0.634)	0.0062 (0.762)	0.0233* (0.078)	0.0058 (0.680)	-0.0115 (0.755)	0.0205 (0.255)	-0.0055 (0.751)
$R^2$	0.376	0.360	0.370	0.302	0.335	0.401	0.358	0.326
Villages	81	79	141	147	154	106	153	154
Households	1115	1038	1154	994	1878	275	1144	1009
$Corr(y, subset)$		-0.064***		0.169***		-0.031		-0.076***
Equal surface coef.	$p = 0.001$		$p = 0.648$		$p = 0.906$		$p = 0.262$	
Equal road coef.	$p = 0.001$		$p = 0.161$		$p = 0.955$		$p = 0.364$	

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

$p$ -values in parentheses, using standard errors clustered at the village level. FDR  $q$ -values in square brackets. All regressions are weighted using person sampling weights from both surveys. Dirt road is the excluded category for the distance variables. The coefficient for "Gravel to Sealed" is not part of the model and comes about by subtracting "Dirt to Gravel" from "Dirt to Sealed". Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications also include location- and household-specific control variables (see Table 4.1.1), province-time-fixed effects, and Mundlak terms. The bottom third row shows correlation coefficients between the outcome variable and subset identifiers. The last two rows contain  $p$ -values of Wald tests for the equality of the coefficients of surface type and of all road variables.

Table 4.3.5: Impact of road type and distances on poverty status by subgroups

	Road distance to nearest town $\leq$ 30km	Road distance to nearest town $>$ 30km	Average years of schooling $\leq$ 4	Average years of schooling $>$ 4	Household head male	Household head female	Household members above 50 $\leq$ 30%	Household members above 50 $>$ 30%
Dirt to Sealed (km)	-0.0082 (0.723) [0.560]	-0.0190*** (0.004) [0.052]	-0.0143 (0.144) [0.230]	-0.0038 (0.677) [0.553]	-0.0162*** (0.003) [0.052]	0.0243 (0.241) [0.279]	-0.0072 (0.371) [0.352]	-0.0192** (0.019) [0.102]
Dirt to Gravel (km)	0.0255* (0.073) [0.189]	-0.0189*** (0.006) [0.052]	-0.0134* (0.073) [0.189]	-0.0023 (0.760) [0.560]	-0.0129** (0.020) [0.102]	0.0026 (0.799) [0.560]	-0.0088 (0.155) [0.232]	-0.0148*** (0.044) [0.174]
Gravel to Sealed (km)	-0.0337** (0.044) [0.174]	-0.0001 (0.975) [0.577]	-0.0009 (0.897) [0.560]	-0.0015 (0.817) [0.560]	-0.0033 (0.323) [0.352]	0.0217 (0.220) [0.265]	0.0016 (0.794) [0.560]	-0.0044 (0.444) [0.361]
Total distance to nearest town (km)	0.0111 (0.163)	0.0004 (0.882)	-0.0007 (0.803)	0.0034 (0.403)	0.0001 (0.944)	-0.0048 (0.365)	0.0016 (0.516)	-0.0021 (0.487)
Distance to nearest road (km)	-0.0356* (0.054)	-0.0024 (0.823)	-0.0025 (0.862)	-0.0164* (0.088)	-0.0074 (0.434)	0.0182 (0.445)	-0.0126 (0.338)	-0.0006 (0.958)
$R^2$	0.261	0.254	0.266	0.210	0.224	0.357	0.239	0.237
Villages	81	79	141	147	154	106	153	154
Households	1115	1038	1154	994	1878	275	1144	1009
$Corr(y, subset)$		0.040*		-0.138***		0.025		0.054***
Equal surface coef.	$p = 0.002$		$p = 0.434$		$p = 0.126$		$p = 0.603$	
Equal road coef.	$p = 0.005$		$p = 0.712$		$p = 0.012$		$p = 0.757$	

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.  
 $p$ -values in parentheses, using standard errors clustered at the village level. FDR  $q$ -values in square brackets. All regressions are weighted using person sampling weights from both surveys. Dirt road is the excluded category for the distance variables. The coefficient for "Gravel to Sealed" is not part of the model and comes about by subtracting "Dirt to Gravel" from "Dirt to Sealed". Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications also include location- and household-specific control variables (see Table 4.1.1), province-time-fixed effects, and Mundlak terms. The bottom third row shows correlation coefficients between the outcome variable and subset identifiers. The last two rows contain  $p$ -values of Wald tests for the equality of the coefficients of surface type and of all road variables.

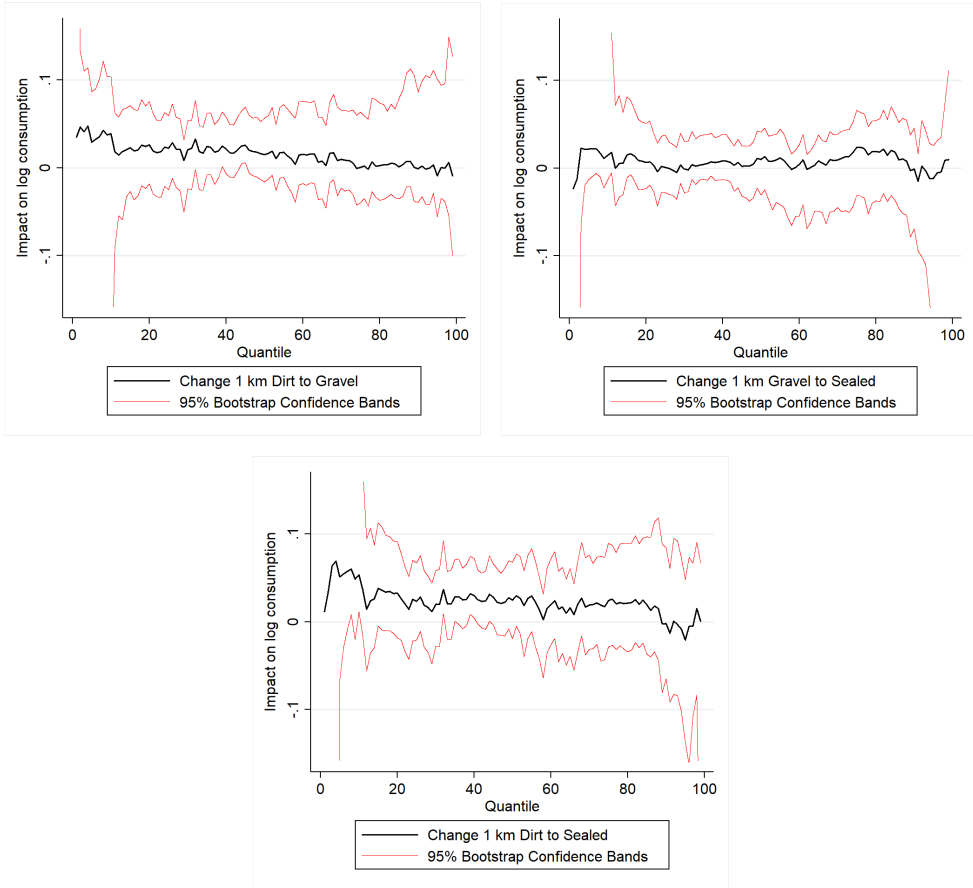


Figure 4.3.1: Road type coefficients by consumption quantile

consequences for material well-being. A short distance to the nearest road, on the other hand, is more relevant to those households living closer to the nearest town. Furthermore, it appears that the likelihood of being poor is less affected by road upgrades for female-led households than for male-led ones, and that households with a comparatively higher share of older household members seem to benefit more from better roads, though the differences in surface coefficients are not jointly significant. Since low consumption and poverty status are positively correlated with remoteness and relatively more older household members, these results also indicate that poorer groups benefit more from better roads.

Last of all, we present the results of the generalized quantile regressions examining the effect of road quality on log per adult-equivalent consumption. Figure 4.3.1 shows

the marginal effects of the difference of one kilometer of road between dirt and gravel, gravel and sealed, and dirt and sealed, for each quantile from 1 to 99. The confidence bands come from our cluster bootstrap. All the control variables from the prior regression models as well as the Mundlak terms are included.

The effect of the difference between a kilometer of sealed and dirt road is on average relatively larger for the lowest 20 quantiles and relatively lower for the highest 20. The average effect on log consumption for the former is 0.041, while it is 0.007 for the latter. Similarly, the effect of the difference between gravel and dirt road is decreasing along the income distribution, with an average effect of 0.030 for the poorest 20 quantiles and 0.001 for the richest 20. However, the large confidence bands generated by the bootstrapping procedure mean that these trends are not statistically significant. We interpret the results from the GQR model as limited evidence that upgrading dirt roads disproportionately benefits the poorest households in PNG and may thus be regarded a pro-poor policy measure.

## 4.4 Conclusion

In this chapter, we examine the effects of the quality of PNG's road network on rural household welfare over a 13-year period. We find evidence that upgrading roads leads to improvements in household welfare, housing quality, and school enrollment. The effects on consumption and poverty are higher for disadvantaged and more remote households. This finding complements the argument by Gibson and Rozelle (2003) that due to the sparse road network and the remoteness of many poor households, infrastructure spending may be one of the few feasible targeted antipoverty measures in PNG. Our results also show that upgrading roads supports the structural transformation of households away from subsistence farming and toward market-oriented activities. This is consistent with the inhomogeneous impact on consumption and suggests that connecting rural households to local markets particularly benefits smallholder farmers.

Our identification strategy makes only use of existing administrative road inventory data in combination with repeated cross-sectional household surveys. These kinds of datasets are available for many other countries. Our method thus easily lends itself for replication elsewhere, and again in PNG after the next national household survey has been conducted.

## 4.A Appendix

### 4.A.1 Road Summary Statistics

Table 4.A.1 shows the road lengths and conditions of PNG’s road network for 2000 and 2009. Tables 4.A.2 and 4.A.3 show the transitions in surface type and condition between the two years for those segments that are included in both maps. The tables reveal no consistent trend in road development. Considering change in surface type, we observe that the length of roads upgraded (i.e., changes from gravel to sealed surface) was roughly offset by roads that deteriorated (gravel to dirt). The characterization of road condition captured in the 2000 and 2009 maps shows substantial improvement (most notably, road condition improving from poor to fair) alongside decline (mainly from good to fair). Table 4.A.4 is an extension of Table 4.1.1, showing summary statistics of surface type *and* condition for the roads to the nearest town. It is noteworthy that for the average household, the route to the nearest town is better than the average PNG road as shown in Table 4.1.1, signified by the higher shares of sealed and gravel roads.

Table 4.A.1: Extent and surface type/condition of the main PNG road network in 2000 and 2009

Surface	Condition	2000		2009	
		Length (km)	Share	Length (km)	Share
Sealed	Good	911	7.8%	1,799	7.0%
Sealed	Fair	914	7.8%	1,067	4.2%
Sealed	Poor	314	2.7%	371	1.5%
Gravel	Good	2,137	18.3%	1,096	4.3%
Gravel	Fair	1,649	14.1%	7,300	28.6%
Gravel	Poor	4,232	36.3%	5,726	22.4%
Dirt	Good	223	1.9%	166	0.7%
Dirt	Fair	63	0.5%	3,660	14.3%
Dirt	Poor	1,230	10.5%	4,332	17.0%
All	All	11,672	100%	25,517	100%

Table 4.A.2: Transition matrix comparing road segment surface types in 2000 and 2009

	Sealed '09	Gravel '09	Dirt '09	Total
Sealed '00	1,821	226	93	2,139
Gravel '00	683	6,502	832	8,017
Dirt '00	27	304	1,185	1,516
Total	2,531	7,031	2,110	11,672

Reported in km. Only listed for stretches where surface type is known in 2000.

Table 4.A.3: Transition matrix comparing road segment conditions in 2000 and 2009

	Good '09	Fair '09	Poor '09	Total
Good '00	1,077	1,531	662	3,270
Fair '00	925	970	731	2,626
Poor '00	457	2,994	2,326	5,776
Total	2,458	5,495	3,719	11,672

Reported in km. Only listed for stretches where road condition is known in 2000.

Table 4.A.4: Summary statistics of road surface type and condition for the analysis sample

	1996			2009/10			<i>p</i> -value	
	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>	<i>N</i>		
<i>Impact variables (obtained from road maps)</i>								
Road to town: km of good sealed road	7.907	12.61	680	7.740	16.35	1,643	0.949	
Road to town: km of fair sealed road	6.834	13.16	680	5.900	15.70	1,643	0.717	
Road to town: km of poor sealed road	1.355	1.962	680	1.041	3.660	1,643	0.504	
Road to town: km of good gravel road	7.315	8.773	680	2.792	8.045	1,643	0.004	
Road to town: km of fair gravel road	10.53	12.78	680	11.09	18.36	1,643	0.830	
Road to town: km of poor gravel road	12.62	17.20	680	3.801	10.82	1,643	0.009	
Road to town: km of good dirt road	0.029	0.205	680	0	0	1,643	0.304	
Road to town: km of fair dirt road	1.696	3.282	680	2.934	8.938	1,643	0.223	
Road to town: km of poor dirt road	1.703	4.707	680	1.265	5.815	1,643	0.626	

Road summary statistics are obtained using household sampling weights. Standard deviations and *p*-values account for clustering at the village level. The *p*-values show the probability of equal means.

#### 4.A.2 Impact Analysis of Surface Type and Condition

Beside the regressions on road surface type, we also consider a more detailed set of specifications that consider surface type *and* road condition. This leaves us with nine categories of road segments (all combinations of sealed, gravel, dirt, and good, fair, poor), some of which are empty or have very few observations. To avoid including variables with little variation, we lump together road categories that accounted for less than 5% of total length. Accordingly, we combine fair and poor sealed roads into one category, and combine good, fair, and poor dirt roads into another category. The excluded category for the regressions is dirt road. The results are summarized in Tables 4.A.5 to 4.A.7, which are the equivalents of Tables 4.3.1 to 4.3.3 in the main text. Conditional on surface type, we see no significant differences in effects by road condition. This does not mean that road condition has no effect. We believe that a combination of insufficient statistical power and quality of the road condition data is responsible for this null result. Road condition can change quickly in PNG, and the way it was assessed likely varied over time and across provinces, making it a variable more prone to measurement error than surface type.



Table 4.A.5: Impact of road type on consumption and poverty status using detailed road type variables

	Log(real p.a.e. consumption)			Poverty Status		
	(1)	(2)	(3)	(1)	(2)	(3)
Sealed and good (km)	-0.0065 (0.138) [0.495]	-0.0088** (0.037) [0.366]	0.0144 (0.245) [0.603]	0.0019 (0.456) [0.786]	0.0031 (0.232) [0.533]	-0.0025 (0.731) [0.854]
Sealed and not good (km)	0.0050 (0.287) [0.659]	0.0056 (0.225) [0.533]	0.0196** (0.036) [0.541]	-0.0040 (0.112) [0.450]	-0.0043* (0.085) [0.384]	-0.0085 (0.126) [0.542]
Gravel and good (km)	-0.0031 (0.540) [0.786]	-0.0050 (0.327) [0.741]	0.0042 (0.684) [0.810]	0.0001 (0.960) [1.000]	0.0011 (0.680) [0.985]	-0.0032 (0.613) [0.734]
Gravel and fair (km)	-0.0026 (0.556) [0.786]	-0.0024 (0.559) [0.985]	0.0058 (0.609) [0.734]	0.0014 (0.536) [0.786]	0.0013 (0.558) [0.985]	-0.0037 (0.577) [0.734]
Gravel and poor (km)	-0.0028 (0.436) [0.786]	-0.0050 (0.155) [0.519]	0.0136 (0.162) [0.572]	0.0012 (0.513) [0.786]	0.0022 (0.225) [0.533]	-0.0069 (0.224) [0.597]
Total distance to nearest town (km)	-0.0021 (0.536) [0.873]	-0.0018 (0.582) [0.454]	0.0017 (0.689) [0.797]	0.0020 (0.236) [0.575]	0.0019 (0.264) [0.258]	0.0005 (0.816) [0.453]
Distance to nearest road (km)	0.0029 (0.873)	0.0119 (0.454)	0.0039 (0.797)	-0.0059 (0.575)	-0.0108 (0.258)	-0.0075 (0.453)
Household controls	No	Yes	Yes	No	Yes	Yes
Mundlak terms included	No	No	Yes	No	No	Yes
$R^2$	0.192	0.323	0.332	0.135	0.218	0.222
Villages	155	155	155	155	155	155
Households	2153	2153	2153	2153	2153	2153
$p$ -value Mundlak			0.003			0.009

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

$p$ -values in parentheses, using standard errors clustered at the village level. FDR  $q$ -values in square brackets. All regressions are weighted using person sampling weights from both surveys. Dirt road is the excluded category for the distance variables. Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications also include location-specific control variables (see Table 4.1.1) as well as province-time-fixed effects. Columns labeled (2) in addition include the household-level control variables (see Table 4.1.1), and columns labeled (3) include Mundlak terms and household-level control variables. The last row displays the  $p$ -value of a Wald test of the joint significance of the Mundlak terms.

Table 4.A.6: Impact of road type on subsistence farming and wage employment using detailed road type variables

	Subsistence farming household			Someone in household has wage job		
	(1)	(2)	(3)	(1)	(2)	(3)
Sealed and good (km)	0.0014 (0.108)	0.0018* (0.071)	-0.0047 (0.183)	0.0016 (0.360)	-0.0003 (0.862)	-0.0074 (0.305)
	[0.450]	[0.384]	[0.597]	[0.762]	[1.000]	[0.604]
Sealed and not good (km)	-0.0000 (0.996)	-0.0001 (0.957)	-0.0033 (0.354)	0.0007 (0.707)	0.0000 (0.979)	-0.0118* (0.058)
	[1.000]	[1.000]	[0.604]	[0.960]	[1.000]	[0.541]
Gravel and good (km)	0.0022* (0.079)	0.0027* (0.056)	0.0023 (0.521)	0.0031* (0.080)	0.0019 (0.200)	0.0014 (0.813)
	[0.403]	[0.383]	[0.734]	[0.403]	[0.533]	[0.910]
Gravel and fair (km)	-0.0007 (0.504)	-0.0005 (0.614)	0.0022 (0.540)	0.0004 (0.800)	0.0001 (0.972)	0.0008 (0.870)
	[0.786]	[0.985]	[0.734]	[1.000]	[1.000]	[0.926]
Gravel and poor (km)	-0.0001 (0.952)	0.0003 (0.754)	0.0009 (0.778)	0.0026 (0.159)	0.0010 (0.523)	0.0024 (0.612)
	[1.000]	[1.000]	[0.895]	[0.495]	[0.985]	[0.734]
Total distance to nearest town (km)	-0.0002 (0.817)	-0.0004 (0.651)	-0.0006 (0.572)	-0.0026** (0.021)	-0.0015 (0.117)	-0.0014 (0.352)
	0.0078* (0.093)	0.0077* (0.085)	0.0079* (0.098)	-0.0365*** (0.000)	-0.0313*** (0.000)	-0.0358*** (0.000)
Distance to nearest road (km)						
Household controls	No	Yes	Yes	No	Yes	Yes
Mundlak terms included	No	No	Yes	No	No	Yes
R <sup>2</sup>	0.150	0.195	0.202	0.155	0.226	0.236
Villages	155	155	155	155	155	155
Households	2323	2322	2322	2312	2311	2311
p-value Mundlak			0.004			0.800

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

p-values in parentheses, using standard errors clustered at the village level. FDR q-values in square brackets. All regressions are weighted using household sampling weights from both surveys. Dirt road is the excluded category for the distance variables. Road sections observed in the 2009 maps, but not in 2000, are assumed to remain in the same over time. All specifications also include location-specific control variables (see Table 4.1.1) as well as province-time-fixed effects. Columns labeled (2) in addition include the household-level control variables (see Table 4.1.1), and columns labeled (3) include Mundlak terms and household-level control variables. The last row displays the p-value of a Wald test of the joint significance of the Mundlak terms.

Table 4.A.7: Impact of road type on having a good roof and school enrollment using detailed road type variables

	Home has a good roof			Ratio of children going to school		
	(1)	(2)	(3)	(1)	(2)	(3)
Sealed and good (km)	0.0033** (0.041) [0.287]	0.0026 (0.104) [0.402]	0.0064 (0.463) [0.734]	0.0031 (0.247) [0.659]	0.0006 (0.802) [1.000]	0.0163* (0.073) [0.541]
Sealed and not good (km)	0.0055** (0.022) [0.274]	0.0052** (0.028) [0.366]	0.0122 (0.132) [0.542]	0.0012 (0.627) [0.824]	0.0012 (0.611) [0.985]	0.0144* (0.087) [0.541]
Gravel and good (km)	0.0047** (0.036) [0.287]	0.0043* (0.057) [0.383]	0.0063 (0.292) [0.604]	0.0046 (0.152) [0.495]	0.0028 (0.344) [0.741]	0.0206** (0.023) [0.541]
Gravel and fair (km)	0.0043** (0.018) [0.274]	0.0042** (0.021) [0.366]	0.0087 (0.138) [0.542]	-0.0015 (0.562) [0.786]	-0.0022 (0.342) [0.741]	0.0135 (0.132) [0.542]
Gravel and poor (km)	0.0063*** (0.000) [0.010]	0.0056*** (0.001) [0.040]	0.0134** (0.015) [0.541]	0.0014 (0.599) [0.817]	-0.0004 (0.885) [1.000]	0.0145* (0.083) [0.541]
Total distance to nearest town (km)	-0.0049*** (0.001) [0.001]	-0.0044*** (0.004) [0.081]	-0.0036* (0.081)	-0.0013 (0.569)	-0.0001 (0.975)	0.0023 (0.186)
Distance to nearest road (km)	-0.0301*** (0.000)	-0.0267*** (0.000)	-0.0284*** (0.000)	-0.0166* (0.072)	-0.0095 (0.233)	-0.0182** (0.050)
Household controls	No	Yes	Yes	No	Yes	Yes
Mundlak terms included	No	No	Yes	No	No	Yes
$R^2$	0.455	0.470	0.474	0.188	0.248	0.256
Villages	155	155	155	154	154	154
Households	2322	2321	2321	1538	1537	1537
$p$ -value Mundlak			0.538			0.188

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

$p$ -values in parentheses, using standard errors clustered at the village level. FDR  $q$ -values in square brackets. The regressions of having a good roof are weighted using household sampling weights from both surveys, the ones of schooling use child sampling weights. Dirt road is the excluded category for the distance variables. Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. All specifications also include location-specific control variables (see Table 4.1.1) as well as province-time-fixed effects. Columns labeled (2) in addition include the household-level control variables (see Table 4.1.1), and columns labeled (3) include Mundlak terms and household-level control variables. The last row displays the  $p$ -value of a Wald test of the joint significance of the Mundlak terms.

### 4.A.3 Selective Migration

To check whether better roads and migration within PNG are correlated, we regress two indicators of migration on our road variables. The first variable is the ratio of prime-aged men in a household, as they are the people most likely to migrate (Aggarwal, 2018). The second variable, which is only available for the HIES 09/10, is a dummy indicating whether anyone among the household head, the spouse, the head's parents, or grandparents, has moved to the current province within the last 10 years. We include location-specific covariates and province-time fixed effects, but we do not include the Mundlak terms, as this is not a causal analysis. Table 4.A.8 shows the regression results. It appears that road quality is statistically insignificant for both migration indicators.

Table 4.A.8: Regressions of indicators of migration on road type and distances

	Ratio of men between 18 and 40	Household member moved to province $\leq$ 10 years ago
Dirt to Sealed (km)	0.0001 (0.773)	-0.0018 (0.156)
Dirt to Gravel (km)	0.0001 (0.864)	-0.0019 (0.104)
Total distance to nearest town (km)	-0.0001 (0.729)	0.0013 (0.124)
Distance to nearest road (km)	0.0013 (0.537)	-0.0034 (0.694)
Household controls	No	No
Mundlak terms included	No	No
$R^2$	0.033	0.043
Villages	155	95
Households	2323	1643
$p$ -value joint significance	0.958	0.231

\* Significant at the 10% level. \*\* Significant at the 5% level. \*\*\* Significant at the 1% level.

$p$ -values in parentheses, using standard errors clustered at the village level. Both regressions are weighted using household sampling weights from both surveys. Dirt road is the excluded category for the distance variables. Road sections observed in the 2009 maps, but not in 2000, are assumed to remain the same over time. Both specifications also include location-specific control variables (see Table 4.1.1) as well as province-time-fixed effects. The last row displays the  $p$ -value of a Wald test of the joint significance of all road variables.

# Bibliography

- Adukia, Anjali, Sam Asher, and Paul Novosad.** 2020. “Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction.” *American Economic Journal: Applied Economics*, 12(1): 348–376.
- Aggarwal, Shilpa.** 2018. “Do Rural Roads Create Pathways out of Poverty? Evidence from India.” *Journal of Development Economics*, 133: 375–395.
- Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, and Julia Tobias.** 2012. “Targeting the Poor: Evidence from a Field Experiment in Indonesia.” *American Economic Review*, 102(4): 1206–1240.
- Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, Ririn Purnamasari, and Matthew Wai-Poi.** 2019. “Does Elite Capture Matter? Local Elites and Targeted Welfare Programs in Indonesia.” *AEA Papers and Proceedings*, 109: 334–339.
- Alatas, Vivi, Ririn Purnamasari, Matthew Wai-Poi, Abhijit Banerjee, Benjamin A Olken, and Rema Hanna.** 2016. “Self-targeting: Evidence from a Field Experiment in Indonesia.” *Journal of Political Economy*, 124(2): 371–427.
- Alkire, Sabina, and James Foster.** 2011. “Counting and Multidimensional Poverty Measurement.” *Journal of Public Economics*, 95(7): 476–487.
- Anderson, Michael L.** 2008. “Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *Journal of the American Statistical Association*, 103(484): 1481–1495.
- Angelucci, Manuela, and Giacomo De Giorgi.** 2009. “Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles’ Consumption?” *American Economic Review*, 99(1): 486–508.

- Angelucci, Manuela, Giacomo De Giorgi, Marcos A. Rangel, and Imran Rasul.** 2010. "Family Networks and School Enrolment: Evidence from a Randomized Social Experiment." *Journal of Public Economics*, 94(3): 197–221.
- Asher, Sam, and Paul Novosad.** 2020. "Rural Roads and Local Economic Development." *American Economic Review*, 110(3): 797–823.
- Athey, Susan, Guido W. Imbens, and Stefan Wager.** 2018. "Approximate Residual Balancing: Debiased Inference of Average Treatment Effects in High Dimensions." *Journal of the Royal Statistical Society Series B*, 80(4): 597–623.
- Attanasio, Orazio P., Costas Meghir, and Ana Santiago.** 2012. "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA." *The Review of Economic Studies*, 79(1): 37–66.
- Bah, Adama, Samuel Bazzi, Sudarno Sumarto, and Julia Tobias.** 2019. "Finding the Poor vs. Measuring Their Poverty: Exploring the Drivers of Targeting Effectiveness in Indonesia." *The World Bank Economic Review*, 33(3): 573–597.
- Banerjee, Abhijit, Esther Duflo, and Nancy Qian.** 2020. "On the Road: Access to Transportation Infrastructure and Economic Growth in China." *Journal of Development Economics*, 145(102442): 1–36.
- Banerjee, Abhijit, Esther Duflo, Raghavendra Chattopadhyay, and Jeremy Shapiro.** 2009. "Targeting Efficiency: How Well Can We Identify the Poorest of the Poor?" *Institute for Financial Management and Research Centre for Micro Finance Working Paper*, 21.
- Bang, Heejung, and James M. Robins.** 2005. "Doubly Robust Estimation in Missing Data and Causal Inference Models." *Biometrics*, 61(4): 962–973.
- Bastagli, Francesca, Jessica Hagen-Zanker, Luke Harman, Valentina Barca, Georgina Sturge, Tanja Schmidt, and Luca Pellerano.** 2016. "Cash Transfers: What Does the Evidence Say? A Rigorous Review of Programme Impact and of the Role of Design and Implementation Features." Overseas Development Institute ODS Report.
- Beggs, Steven, Scott Cardell, and Jerry A. Hausman.** 1981. "Assessing the Potential Demand for Electric Cars." *Journal of Econometrics*, 17(1): 1–19.

- Behrman, Jere R., and Petra E Todd.** 1999. "Randomness in the Experimental Samples of PROGRESA (Education, Health, and Nutrition Program)." *International Food Policy Research Institute, Washington, DC*.
- Behrman, Jere R., Piyali Sengupta, and Petra Todd.** 2005. "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico." *Economic Development and Cultural Change*, 54(1): 237–275.
- Behrman, Jere R., Susan W Parker, and Petra E Todd.** 2009. "7 Medium-Term Impacts of the Oportunidades Conditional Cash Transfer Program on Rural Youth in Mexico." *Poverty, Inequality, and Policy in Latin America*, 219.
- Bell, Clive, and Susanne van Dillen.** 2020. "How does India's Rural Roads Program Affect the Grassroots? Findings from a Survey in Orissa." *Land Economics*, 90(2): 372–394.
- Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen.** 2014. "Inference on Treatment Effects after Selection among High-Dimensional Controls." *The Review of Economic Studies*, 81(2): 608–650.
- Belloni, Alexandre, Victor Chernozhukov, Ivan Fernández-Val, and Christian Hansen.** 2017. "Program Evaluation and Causal Inference With High-Dimensional Data." *Econometrica*, 85(1): 233–298.
- Binswanger, Hans P., Shahidur R. Khandker, and Mark R. Rosenzweig.** 1993. "How Infrastructure and Financial Institutions Affect Agricultural Output and Investment in India." *Journal of Development Economics*, 41(2): 337–366.
- Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel.** 2015. "The Value of Democracy: Evidence from Road Building in Kenya." *American Economic Review*, 105(6): 1817–1851.
- Buser, Thomas, Hessel Oosterbeek, Erik Plug, Juan Ponce, and José Rosero.** 2017. "The Impact of Positive and Negative Income Changes on the Height and Weight of Young Children." *World Bank Economic Review*, 31(3): 786–808.
- Cameron, Lisa, and Manisha Shah.** 2014. "Can Mistargeting Destroy Social Capital and Stimulate Crime? Evidence from a Cash Transfer Program in Indonesia." *Economic Development and Cultural Change*, 62(2): 381–415.

- Carrell, Scott E., and Mark L. Hoekstra.** 2010. "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids." *American Economic Journal: Applied Economics*, 2(1): 211–228.
- Casaburi, Lorenzo, Rachel Glennerster, and Tavneet Suri.** 2013. "Rural Roads and Intermediated Trade: Regression Discontinuity Evidence from Sierra Leone." *Working Paper*.
- Castaing Gachassin, Marie.** 2013. "Should I Stay or Should I Go? The Role of Roads in Migration Decisions." *Journal of African Economics*, 22(5): 796–826.
- Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins.** 2018. "Double/Debiased Machine Learning for Treatment and Structural Parameters." *The Econometrics Journal*, 21(1): 1–68.
- Coady, David, Margaret Grosh, and John Hoddinott.** 2004. "Targeting Outcomes Redux." *The World Bank Research Observer*, 19(1): 61–85.
- Coady, David P., and Susan W. Parker.** 2009. "Targeting Performance under Self-selection and Administrative Targeting Methods." *Economic Development and Cultural Change*, 57(3): 559–587.
- Crump, Richard K., V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik.** 2009. "Dealing with Limited Overlap in Estimation of Average Treatment Effects." *Biometrika*, 96(1): 187–199.
- Deci, Edward, Richard Koestner, and Richard Ryan.** 1999. "A Meta-Analytic Review of Experiments Examining the Effect of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin*, 125: 627–668.
- de Janvry, Alain, Frederico Finan, Elisabeth Sadoulet, and Renos Vakis.** 2006. "Can Conditional Cash Transfer Programs Serve as Safety Nets in Keeping Children at School and from Working When Exposed to Shocks?" *Journal of Development Economics*, 79(2): 349–373. Special Issue in Honor of Pranab Bardhan.
- Dercon, Stefan, Daniel O. Gilligan, John Hoddinott, and Tassew Woldehanna.** 2009. "The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages." *American Journal of Agricultural Economics*, 91(4): 1007–1021.



- Dercon, Stefan, John Hoddinott, and Tassew Woldehanna.** 2012. "Growth and Chronic Poverty: Evidence from Rural Communities in Ethiopia." *The Journal of Development Studies*, 48(2): 238–253.
- Dizon-Ross, Rebecca.** 2018. "Parents' Beliefs About Their Children's Academic Ability: Implications for Educational Investments." National Bureau of Economic Research Working Paper 24610.
- Dornan, Matthew.** 2016. "The Political Economy of Road Management Reform: Papua New Guinea's National Road Fund: Road Management Reform." *Asia & the Pacific Policy Studies*, 3: 443–457.
- Dubois, Pierre, Alain de Janvry, and Elisabeth Sadoulet.** 2012. "Effects on School Enrollment and Performance of a Conditional Cash Transfer Program in Mexico." *Journal of Labor Economics*, 30(3): 555–589.
- Edmonds, Christopher, Martin Wiegand, Eric Koomen, Menno Pradhan, and Bo Andree.** 2018. "Assessing the Impact of Road Development on Household Welfare in Rural Papua New Guinea." In *Financing Infrastructure in Asia: Capturing Impacts and New Sources*. , ed. Naoyuki Yoshino, Matthias Helble and Umid Abidhadjaev, Chapter 7, 189–235. Asian Development Bank Institute.
- Farrell, Max H.** 2015. "Robust Inference on Average Treatment Effects with Possibly More Covariates than Observations." *Journal of Econometrics*, 189(1): 1–23.
- Fehr, Ernst, and Armin Falk.** 2002. "Psychological Foundations of Incentives." Institute for Empirical Research in Economics - University of Zurich IEW - Working Papers 095.
- Fiszbein, Ariel, and Norbert R. Schady.** 2010. *Conditional Cash Transfers*. The World Bank.
- Frey, Bruno S., and Reto Jegen.** 2001. "Motivation Crowding Theory." *Journal of Economic Surveys*, 15(5): 589–611.
- Gantner, Leigh.** 2009. "PROGRESA: An Integrated Approach to Poverty Alleviation in Mexico." In *Case Studies in Food Policy for Developing Countries: Policies for Health, Nutrition, Food Consumption, and Poverty*. , ed. Per Pinstrup-Andersen and Fuzhi Cheng, 211–220. Cornell University Press.

- Gertler, Paul J., Marco Gonzalez-Navarro, Tadeja Gracner, and Alexander D. Rothenberg.** 2016. "Road Quality, Local Economic Activity, and Welfare: Evidence from Indonesia's Highways." *CEGA Working Paper Series No. WPS-058*, Center for Effective Global Action. University of California, Berkeley.
- Gibson, John.** 2012. "Papua New Guinea Poverty Profile. Based on the Household Income and Expenditure Survey." Technical report.
- Gibson, John, and Scott Rozelle.** 1998. "Results of the Household Survey Component of the 1996 Poverty Assessment for Papua New Guinea." Technical report.
- Gibson, John, and Scott Rozelle.** 2003. "Poverty and Access to Roads in Papua New Guinea." *Economic Development and Cultural Change*, 52(1): 159–185.
- Gibson, John, and Susan Olivia.** 2010. "The Effect of Infrastructure Access and Quality on Non-Farm Enterprises in Rural Indonesia." *World Development*, 38(5): 717–726.
- Gneezy, Uri, and Aldo Rustichini.** 2000. "A Fine Is a Price." *The Journal of Legal Studies*, 29(1): 1–17.
- Hurvich, Clifford M., and Chih-Ling Tsai.** 1989. "Regression and Time Series Model Selection in Small Samples." *Biometrika*, 76(2): 297–307.
- Imbens, Guido W., and Donald B. Rubin.** 2015. "Trimming to Improve Balance in Covariate Distributions." *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, 359–374. Cambridge University Press.
- Jensen, Robert.** 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *The Quarterly Journal of Economics*, 125(2): 515–548.
- Jusi, Petri, Roy Mumu, Sirpa H. Jarvenpaa, Barnabas Neausemale, and Jr. Eduardo Sangrador.** 2003. "Road Asset Management System Implementation in Pacific Region: Papua New Guinea." *Transportation Research Record*, 1819(1): 323–332.
- Kahneman, Daniel, and Amos Tversky.** 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47(2): 263–291.
- Kanbur, Ravi, and Paul Shaffer.** 2007. "Epistemology, Normative Theory and Poverty Analysis: Implications for Q-Squared in Practice." *World Development*, 35(2): 183–196. Experiences of Combining Qualitative and Quantitative Approaches in Poverty Analysis.

- Karlan, Dean, and Bram Thuysbaert.** 2019. "Targeting Ultra-Poor Households in Honduras and Peru." *The World Bank Economic Review*, 33(1): 63–94.
- Khandker, Shahidur R., and Gayatri B. Koolwal.** 2010. "How Infrastructure and Financial Institutions Affect Rural Income and Poverty: Evidence from Bangladesh." *The Journal of Development Studies*, 46(6): 1109–1137. PMID: 20645462.
- Khandker, Shahidur R., and Gayatri B. Koolwal.** 2011. "Estimating the Long-Term Impacts of Rural Roads: A Dynamic Panel Approach." *Policy Research working paper. WPS 5867.*, Washington, DC: World Bank.
- Khandker, Shahidur R., Zaid Bakht, and Gayatri B. Koolwal.** 2009. "The Poverty Impact of Rural Roads: Evidence from Bangladesh." *Economic Development and Cultural Change*, 57(4): 685–722.
- Kidd, Stephen, and Diloá Athias.** 2020. "Hit and Miss: An Assessment of Targeting Effectiveness in Social Protection." Development Pathways and Church of Sweden.
- Knox, Jerry, Andre Daccache, and Tim Hess.** 2013. "What is the Impact of Infrastructural Investments in Roads, Electricity and Irrigation on Agricultural Productivity?" CEE review 11-007. Collaboration for Environmental Evidence (CEE) Syntheses.
- Kwa, Eric, Stephen Howes, and Soe Lin.** 2010. "Review of the PNG-Australia Development Cooperation Treaty (1999)." Department of Foreign Affairs and Trade, Australian Government.
- Lavy, Victor, M. Daniele Paserman, and Analta Schlosser.** 2012. "Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom." *The Economic Journal*, 122(559): 208–237.
- Loewenstein, George F.** 1988. "Frames of Mind in Intertemporal Choice." *Management Science*, 34(2): 200–214.
- Lokshin, Michael, and Ruslan Yemtsov.** 2005. "Has Rural Infrastructure Rehabilitation in Georgia Helped the Poor?" *World Bank Economic Review*, 19: 311–333.
- Lucius, D.** 2010. "Civil Works Capacity Constraints." Asian Development Bank, Port Moresby.
- Mackinnon, James G., and Matthew D. Webb.** 2017. "Wild Bootstrap Inference for Wildly Different Cluster Sizes." *Journal of Applied Econometrics*, 32(2): 233–254.

- Mammen, Enno.** 1993. "Bootstrap and Wild Bootstrap for High Dimensional Linear Models." *The Annals of Statistics*, 21(1): 255–285.
- Mansuri, Ghazala, and Vijayendra Rao.** 2013. *Localizing Development: Does Participation Work?* Washington, D.C.:The World Bank.
- Mundlak, Yair.** 1978. "On the Pooling of Time Series and Cross Section Data." *Econometrica*, 46(1): 69–85.
- Mu, Ren, and Dominique van de Walle.** 2011. "Rural Roads and Local Market Development in Vietnam." *The Journal of Development Studies*, 47(5): 709–734.
- Nguyen, Cuong Viet.** 2019. "Impacts of Rural Roads on Household Welfare in Vietnam: Evidence from a Replication Study." *Journal of Economics and Development*, 21(1): 83–112.
- Nguyen, Trang.** 2008. "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar." *Unpublished manuscript*.
- Olson Lanjouw, Jean, and Peter Lanjouw.** 2001. "How to Compare Apples and Oranges: Poverty Measurement Based on Different Definitions of Consumption." *Review of Income and Wealth*, 47(1): 25–42.
- Oster, Emily.** 2017. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics*, 1–18.
- Powell, David.** 2020. "Quantile Treatment Effects in the Presence of Covariates." *The Review of Economics and Statistics*, 102(5): 994–1005.
- Ravallion, Martin.** 2009. "How Relevant Is Targeting to the Success of an Antipoverty Program?" *The World Bank Research Observer*, 24(2): 205–231.
- Rosenbaum, Paul R., and Donald B. Rubin.** 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, 70(1): 41–55.
- Schultz, T. Paul.** 2004. "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program." *Journal of Development Economics*, 74(1): 199–250. New Research on Education in Developing Economies.
- Schüring, Esther.** 2014. "Preferences for Community-based Targeting – Field Experimental Evidence from Zambia." *World Development*, 54: 360–373.

- Sen, Amartya.** 1980. "Equality of What?" *The Tanner Lecture on Human Values*, I: 197–220.
- Shaffer, Paul.** 2013. "Ten Years of "Q-Squared": Implications for Understanding and Explaining Poverty." *World Development*, 45: 269–285.
- Skoufias, Emmanuel.** 2005. "PROGRESA and Its Impacts on the Welfare of Rural Households in Mexico." Research Report of the International Food Policy Research Institute.
- Slattery, David, Matthew Dornan, and John Lee.** 2018. "Road Management in Papua New Guinea: An Evaluation of a Decade of Australian Support 2007-2017." Australian Government Department of Foreign Affairs and Trade, Canberra.
- Stead, Doug.** 1990. "Engineering Geology in Papua New Guinea: A Review." *Engineering Geology*, 29(1): 1–29.
- Thaler, Richard.** 1980. "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior & Organization*, 1(1): 39–60.
- Todd, Petra E., and Kenneth I. Wolpin.** 2006. "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility." *The American Economic Review*, 96(5): 1384–1417.
- Tversky, Amos, and Daniel Kahneman.** 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science*, 185(4157): 1124–1131.
- Tversky, Amos, and Daniel Kahneman.** 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *The Quarterly Journal of Economics*, 106(4): 1039–1061.
- Weiss, Yoram, Arden Hall, and Fred Dong.** 1980. "The Effect of Price and Income on Investment in Schooling." *Journal of Human Resources*, 15(4): 611–640.
- Widjaja, Muliadi.** 2012. "An Economic and Social Review on Indonesia's Direct Cash Transfer Program to Poor Families in 2005." *Economics and Finance in Indonesia*, 60: 197–212.
- World Food Programme and Logistics Cluster.** 2011. "Papua New Guinea Emergency Preparedness: Operational Logistics Contingency Plan Part 2 – Existing Response Capacity and Overview of Logistics Situation."

**Yusuf, Moeed.** 2010. *Community Targeting for Poverty Reduction: Lessons from Developing Countries*. Boston University.

**Zeller, Manfred, Joseph Feulefack, and Andreas Neef.** 2006. "How Accurate is Participatory Wealth Ranking (PWR) in Targeting the Poor? A Case Study from Bangladesh." International Association of Agricultural Economists 2006 Annual Meeting, August 12-18, 2006, Queensland, Australia.

# Summary

The three main chapters in this dissertation cover three different topics in development economics: education, poverty, and infrastructure. Each chapter contains both relevant findings for the respective subject area and methodological contributions.

In chapter 2, I study the impact of a conditional cash transfer (CCT) program on students' decisions to continue school once the program ends. The CCT in question is Mexico's PROGRESA, which covered students only until the end of middle school when it was introduced. I find that the program reduces the probability to transfer to high school afterwards by about 10 to 14 percentage points. After ruling out competing explanations, the reasons appear to be behavioral: cash crowds out the intrinsic motivation for seeking education, and CCT programs ending early may signal that school is not worth it after a certain point. In addition, the program had positive spillover effects on those students who were not eligible. The program seems to have raised the desire of students from better-off families to distinguish themselves by choosing to stay in school. Identifying the transition probability poses a unique challenge: if PROGRESA has successfully kept students in middle school, then samples of middle school graduates who received the program are likely different from those who did not. To tackle this issue, I compare students by pre-treatment characteristics. I apply *double machine learning*—a recently developed method to identify treatment effects in the presence of a large number of potential confounders. I extend the method to account not only for selection bias but also for non-random attrition. This approach may prove useful in general for studying the aftereffects of programs that impact the composition of its participants.

Chapter 3 is about participatory wealth rankings (PWRs)—a targeting method in which representatives of a community rank households by their wealth. I demonstrate how PWRs can be used to construct a welfare measure that reflects local perceptions of poverty, using data from a field experiment on poverty targeting in Indonesia. The idea is to estimate the relationship between rankings and household characteristics, via a rank-ordered logit model, and to predict welfare scores. This makes it possible to compare households from different villages, even though they have not been compared in the

PWRs. I then estimate the impact of using this new welfare measure as targeting goal on program satisfaction of villagers and village leaders. I find that higher targeting accuracy, as measured by the rank-based welfare measure, significantly increases satisfaction with an anti-poverty program. Furthermore, after controlling for targeting accuracy, the PWR does not lead to significantly higher satisfaction than consumption-based targeting methods. Lastly, I find that targeting accuracy explains satisfaction outcomes better when it is measured against rank-based welfare measure rather than predicted consumption. This holds true even for communities where no PWRs had been conducted. Taken together, the results show that taking local welfare perceptions into account leads to more satisfactory targeting outcomes, while the actual procedure of meeting and ranking households seems to matter little, if at all. This insight is useful especially for contexts in which PWRs are not feasible.

In chapter 4, which is joint work with Eric Koomen, Menno Pradhan, and Christopher Edmonds, we study the impact of road development on household welfare in rural Papua New Guinea (PNG). Using two household surveys from 1996 and 2010 as well as corresponding road maps, we construct road quality variables for the route connecting households to urban areas. We use a correlated random effects model to account for unobserved location-specific effects that might influence both road quality and households' well-being. Our results show that upgrading the roads leading to the nearest town increases average household consumption, housing quality, and school enrollment, and reduces reliance on subsistence farming. An analysis by subgroups shows that the effects on consumption and poverty are at least twice as high for households with a road distance of at least 30 km to the nearest town when comparing them to those living closer than 30 km. Furthermore, we apply a newly developed generalized quantile regression estimator to look for effect heterogeneity along the distribution of consumption. The estimates indicate that upgrading dirt roads has a higher effect for the poorest households. Taken together, these results imply that road infrastructure programs may be considered pro-poor policy measures.







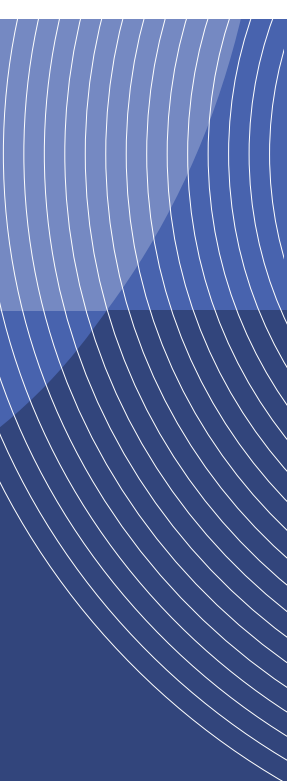
The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

- 736 S. SÓVÁGÓ, *Where to Go Next? Essays on the Economics of School Choice*
- 737 M. HENNEQUIN, *Expectations and Bubbles in Asset Market Experiments*
- 738 M.W. ADLER, *The Economics of Roads: Congestion, Public Transit and Accident Management*
- 739 R.J. DÖTTLING, *Essays in Financial Economics*
- 740 E.S. ZWIERS, *About Family and Fate: Childhood Circumstances and Human Capital Formation*
- 741 Y.M. KUTLUAY, *The Value of (Avoiding) Malaria*
- 742 A. BOROWSKA, *Methods for Accurate and Efficient Bayesian Analysis of Time Series*
- 743 B. HU, *The Amazon Business Model, the Platform Economy and Executive Compensation: Three Essays in Search Theory*
- 744 R.C. SPERNA WEILAND, *Essays on Macro-Financial Risks*
- 745 P.M. GOLEC, *Essays in Financial Economics*
- 746 M.N. SOUVERIJN, *Incentives at work*
- 747 M.H. COVENEY, *Modern Imperatives: Essays on Education and Health Policy*
- 748 P. VAN BRUGGEN, *On Measuring Preferences*
- 749 M.H.C. NIENTKER, *On the Stability of Stochastic Dynamic Systems and their use in Econometrics*
- 750 S. GARCIA MANDICÓ, *Social Insurance, Labor Supply and Intra-Household Spillovers*
- 751 Y. SUN, *Consumer Search and Quality*
- 752 I. KERKEMEZOS, *On the Dynamics of (Anti) Competitive Behaviour in the Airline Industry*

- 753 G.W. GOY, *Modern Challenges to Monetary Policy*
- 754 A.C. VAN VLODROP, *Essays on Modeling Time-Varying Parameters*
- 755 J. SUN, *Tell Me How To Vote, Understanding the Role of Media in Modern Elections*
- 756 J.H. THIEL, *Competition, Dynamic Pricing and Advice in Frictional Markets: Theory and Evidence from the Dutch Market for Mortgages*
- 757 A. NEGRIU, *On the Economics of Institutions and Technology: a Computational Approach*
- 758 F. GRESNIGT, *Identifying and Predicting Financial Earth Quakes using Hawkes Processes*
- 759 A. EMIRMAHMUTOGLU, *Misperceptions of Uncertainty and Their Applications to Prevention*
- 760 A. RUSU, *Essays in Public Economics*
- 761 M.A. COTOFAN, *Essays in Applied Microeconomics: Non-Monetary Incentives, Skill Formation, and Work Preferences*
- 762 B.P.J. ANDRÉE, *Theory and Application of Dynamic Spatial Time Series Models*
- 763, P. PELZL, *Macro Questions, Micro Data: The Effects of External Shocks on Firms*
- 764 D.M. KUNST, *Essays on Technological Change, Skill Premia and Development*
- 765 A.J. HUMMEL, *Tax Policy in Imperfect Labor Markets*
- 766 T. KLEIN, *Essays in Competition Economics*
- 767 M. VIGH, *Climbing the Socioeconomic Ladder: Essays on Sanitation and Schooling*
- 768 Y. XU, *Eliciting Preferences and Private Information: Tell Me What You Like and What You Think*
- 769 S. RELLSTAB, *Balancing Paid Work and Unpaid Care over the Life-Cycle*
- 770 Z. DENG, *Empirical Studies in Health and Development Economics*
- 771 L. KONG, *Identification Robust Testing in Linear Factor Models*

- 772 I. NEAMȚU, *Unintended Consequences of Post-Crisis Banking Reforms*
- 773 B. KLEIN TEESELINK, *From Mice to Men: Field Studies in Behavioral Economics*
- 774 B. TEREICK, *Making Crowds Wiser: The Role of Incentives, Individual Biases, and Improved Aggregation*
- 775 A. CASTELEIN, *Models for Individual Responses*
- 776 D. KOLESNYK, *Consumer Disclosures on Social Media Platforms: A Global Investigation*
- 777 M.A. ROLA-JANICKA, *Essays on Financial Instability and Political Economy of Regulation*
- 778 J.J. KLINGEN, *Natural Experiments in Environmental and Transport Economics*
- 779 E.M. ARTMANN, *Educational Choices and Family Outcomes*
- 780 F.J. OSTERMEIJER, *Economic Analyses of Cars in the City*
- 781 T. ÖZDEN, *Adaptive Learning and Monetary Policy in DSGE Models*
- 782 D. WANG, *Empirical Studies in Financial Stability and Natural Capital*
- 783 L.S. STEPHAN, *Estimating Diffusion and Adoption Parameters in Networks  
New Estimation Approaches for the Latent-Diffusion-Observed-Adoption Model*
- 784 S.R. MAYER, *Essays in Financial Economics*
- 785 A.R.S. WOERNER, *Behavioral and Financial Change – Essays in Market Design*





This dissertation consists of three chapters on different topics within development economics. Chapter 2 explores the question whether conditional cash transfer programs that end early have adverse effects on continued education. Looking at Mexico's PROGRESA, where payments ended after middle school, the study finds that the program reduced the likelihood for students to continue with high school. Chapter 3 is about participatory wealth rankings—a poverty targeting method in which representatives of a community rank households by their wealth. The study demonstrates how such rankings can be used to construct and utilize a welfare measure that reflects local perceptions of poverty, based on data from a field experiment in Indonesia. Chapter 4 investigates the impact of road development on household welfare in rural Papua New Guinea. A special focus lies on the question whether improving roads can be viewed as a pro-poor policy.

